International Journal of Human-Computer Studies 131 (2019) 99–107

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

# History and future of human-automation interaction

Christian P. Janssen[a,*], Stella F. Donker[a], Duncan P. Brumby[b], Andrew L. Kun[c]

[a] Utrecht University, Experimental Psychology and Helmholtz Institute, Heidelberglaan 1, Utrecht 3584 CS, the Netherlands
[b] UCL Interaction Centre, University College London, London WC1E 6BT, United Kingdom
[c] University of New Hampshire, Electrical and Computer Engineering, Kingsbury Hall, Durham, NH 03824, USA

ARTICLE INFO

*Keywords:*
Automation
Human-automation interaction
Safety-critical systems
Autonomous agents
Embodied systems
Situated systems
Divided attention
Ethics
Robotics, Automated vehicles

ABSTRACT

We review the history of human-automation interaction research, assess its current status and identify future directions. We start by reviewing articles that were published on this topic in the International Journal of Human-Computer Studies during the last 50 years. We find that over the years, automated systems have been used more frequently (1) in time-sensitive or safety-critical settings, (2) in embodied and situated systems, and (3) by non-professional users. Looking to the future, there is a need for human-automation interaction research to focus on (1) issues of function and task allocation between humans and machines, (2) issues of trust, incorrect use, and confusion, (3) the balance between focus, divided attention and attention management, (4) the need for interdisciplinary approaches to cover breadth and depth, (5) regulation and explainability, (6) ethical and social dilemmas, (7) allowing a human and humane experience, and (8) radically different human-automation interaction.

## 1. Introduction

The concepts of automation, and mechanized and automated work have been around for decades. According to the Britannica encyclopedia, automation is "*the application of machines to tasks once performed by human beings or, increasingly, to tasks that would otherwise be impossible. Although the term mechanization is often used to refer to the simple replacement of human labour by machines, automation generally implies the integration of machines into a self-governing system.*" (Groover, 2018).

The above definition of automation does not involve the requirement of a computer processor. However, many modern forms of automated (or sometimes: autonomous) machines, such as power plant monitoring devices, automated cars, drones, robots, and chatbots, do involve computers. These computer-automated systems are used by humans, and humans are expected to remain essential contributors to artificial systems and automated systems in the future (Stone et al., 2016). The study of human-computer interaction, or more specifically human-automation interaction, therefore continues to remain relevant as automated systems are used to support more and more everyday activities, overseen by non-technical and non-professional end-users.

In this special issue to celebrate the 50th anniversary of the International Journal of Human-Computer Studies, and its predecessor the International Journal of Man-Machine Studies (from now on collectively referred to as IJHCS), we review the contributions that IJHCS has made towards the study of human-automation interaction. We therefore analyze published work from the journal to distill historic trends. Our analysis shows that human-automation interaction is a field that keeps expanding into new domains and contexts (what we refer to as "breadth"), and also keeps improving its performance within domains and contexts (what we refer to as "depth"). Given these expansions, and the exposure to more contexts and to a wider and more diverse group of end-users, there is a potential for the broader human-computer interaction community to contribute skills and knowledge to create and evaluate safe, engaging, and productive automated systems.

We close our analysis by discussing eight trends that we deem of particular relevance for this community, classified in two segments. First, we discuss trends that have been around for a while but continue to remain important: (1) function and task allocation between humans and machines, (2) trust, incorrect use, and confusion, and (3) the balance between focus, divided attention and attention management. Then, we discuss emerging themes: (4) the need for interdisciplinary approaches to cover breadth and depth, (5) regulation and explainability, (6) ethical and social dilemmas, (7) allowing a human and humane experience, and (8) radically different human-automation interaction.

* Corresponding author.
*E-mail address:* c.p.janssen@uu.nl (C.P. Janssen).

https://doi.org/10.1016/j.ijhcs.2019.05.006
Received 29 January 2019; Received in revised form 2 May 2019; Accepted 16 May 2019
Available online 17 May 2019
1071-5819/ 

**Table 1**
Articles in IJHCS that self-identified as covering automation, per decade compared to the total number of articles that appeared in the journal that year.

| Topic | 1969–1979 | 1980–1989 | 1990–1999 | 2000–2009 | 2010–2019 | Total |
|---|---|---|---|---|---|---|
| Articles on “automated”, “automation” or “autonomous” in IJHCS | 13 | 82 | 37 | 49 | 37 | **218** |
| Reference: total articles in IJHCS per decade | 334 | 696 | 764 | 728 | 727 | **3249** |

## 2. History of human-automation interaction

To gain an overview of the number of articles that were published on the topic of human-automation interaction in IJHCS over its 50 year existence, we conducted a Scopus search on January 14th 2019. We collected all articles that had the word “automation”, “automated”, or “autonomous” in either the title, abstract, or keywords. Table 1 reports the number of articles that matched the search query per topic and decade, together with the total number of articles that was published in IJHCS that decade.

The topic of automation covers a substantial subset of the published work in IJHCS: 4–11% of published articles in each decade, with around 5–6% of the articles in the last two decades. These percentages should be interpreted as approximate values, as the count is limited by the keywords that authors used in their paper's title, abstract and keywords section. There might be false alarms (papers that were returned based on keywords, but that did not directly address research on human-automation interaction) and misses (papers that are relevant for the field of human-automation interaction, but did not include these specific keywords).

To gain a richer understanding of the themes that are discussed in IJHCS papers on human-automation interaction, our initial keyword search was followed by a qualitative analysis. For this analysis, we sorted the IJHCS papers on human-automation interaction by year of publication. We then read the titles and abstracts of these papers to pick up common themes per decade. This revealed four themes which align well with more general trends in artificial intelligence (e.g., Russell and Norvig, 2009, chapter 1) and human-computer interaction (e.g., Carroll, 2013). However, as the analysis method is subjective in nature, and limited by the papers that were published in IJHCS, we do not claim that we have identified all strands of human-automation interaction research that occurred over the last five decades. We do claim that we identified relevant themes, which are discussed in more detail next.

### 2.1. Start: automation for dedicated domains

Publications on automation in IJHCS largely started off with the study of dedicated, domain specific systems. In the 1970s and 1980s a large proportion of published work (around 25 papers) focused specifically on the development and evaluation of automated psychological tests (for overview papers, see e.g. Elithorn et al., 1982; Thompson and Wilson, 1982). The widespread introduction of computers allowed psychology researchers to conduct interactive tasks on computers, instead of just pen-and-paper tests or subjective assessment. Nowadays, digital testing is common in experimental studies involving human participants, and has given rise to opportunities for conducting large-scale studies using crowdsourcing platforms, like Amazon's Mechanical Turk (see Gould et al., 2018 for a review). Given the rise and ubiquity of personal computing devices, the idea of completing an online survey would now hardly qualify as an example of “automation” anymore.

A second dedicated domain in which automation was researched is knowledge acquisition (Feigenbaum, 1977). As reviewed in a previous IJHCS special issue (Motta, 2013), one of the main aims within this domain in the 1980s was to be able to develop methods to ‘extract’ knowledge from experts that can be represented in machines. Among our dataset of papers on automation, the top-cited papers from the 1980s all proposed methods for knowledge elicitation (e.g., Belkin et al., 1987; Diederich et al., 1987; Gruber and Cohen, 1987). Since the 1980s there has been a general shift in perspective that successful knowledge acquisition and knowledge engineering requires more than extracting knowledge. Considerations of systems engineering and allowing smart inferences based on multiple sources (e.g., through the internet) are now seen to be key, with modern day knowledge acquisition research taking on a broad and multi-disciplinary perspective (see also Motta, 2013; Gaines, 2013; Breuker, 2013).

### 2.2. Time-sensitive and safety-critical settings

Throughout the last five decades of IJHCS, automation research has branched out into more domains and settings. One distinct class of research is on tasks that are time-sensitive (i.e., require a response within a finite, short time interval) and/or safety-critical (i.e., where an incorrect action can have disastrous consequences). Work in this area has been published in every decade, but particularly in the 1990s and early 2000s. The range of settings in which time-sensitive and safety-critical tasks have been studied is diverse and varied: from monitoring dynamic processes in factories (e.g., Lee and Moray, 1994), power plants (e.g., Vicente et al., 2001), and other professional settings (e.g., Bahner et al., 2008; van Gigh, 1971), to flight monitoring (e.g., Singh et al., 1997; Skitka et al., 1999, 2000), and semi-automated driving (e.g., Rajaonah et al., 2008, Seppelt and Lee, 2007).

The diversity of domains (and the importance of preventing incidents) has allowed an exploration of deep general topics throughout the history of IJHCS, which remain relevant for today's research. They include topics such as how to distribute or allocate tasks between humans and machines (Dearden et al., 2000; Hollnagel and Bye, 2000; Press, 1971; Sheridan, 2000; de Vries et al., 2003; Milewski and Lewis, 1997), finding the right levels of workload to avoid under- and overload (Van Gigh, 1971; Rajaonah et al., 2008), how to promote appropriate levels of trust in automation (Dzindolet et al., 2003; Lee and Moray, 1994), and how to avoid incorrect use and (human) errors such as through complacency (Bahner et al., 2008) or (human) biases (Skitka et al., 1999, 2000). We will return to the current status of these topics in more detail in our section on the future of human-automation interaction.

### 2.3. Embodied, situated agents

Since the 1990s there has been a gradual shift away from static systems for specific domains (e.g., expert systems, systems for psychological testing) to systems that involve a dynamic intelligent agent that performs a task (e.g., Milewski and Lewis, 1997; Zeng and Sycara, 1998). This continues in the 2000s, with a rise of papers on automated systems that act in a dynamic, physical world. This parallels the popularization in Artificial Intelligence (AI) research of embodied, situated agents (Pfeifer and Scheier, 2001): systems that have their own sensors and that depend on interaction with the environment for performance. For example, in the 2000s IJHCS published various studies on physical robots (e.g., Kaber et al., 2006; Sakamoto et al., 2005) and cars (Rajaonah et al., 2008; Seppelt and Lee, 2007). In parallel, there is also research published on affective interaction with robots, and automated (emotion) feature detection (e.g., Bailenson et al., 2008; Brave et al., 2005; Partala and Surakka, 2003). These topics continue in the 2010s, but also broaden out to include, for example, research on human-robot interaction with multiple robots (Chien et al., 2018).

The relevance of considering embodied and situated robotics and automation explicitly is that the actions of embodied, situated systems (at least in part) depend on how the world is perceived through the machine's sensors, and through the environment in which the machine interacts (Pfeifer and Scheier, 2001). Different machines can (learn to) act differently if either their sensors have different capabilities or if they are trained in different kinds of environments.

Generalization to unknown settings, and adaptation to new settings, requires extensive training for these embodied, situated robots. Automated vehicles are an example of an embodied, situated robot that acts in and adapts to unknown settings. For automated vehicles, training typically consists of a combination of extensive experience under real-world driving conditions, as well as extensive simulated training sessions to learn how to act in other potential worlds (Madrigal, 2017). By contrast, earlier simpler automated systems, such as, closed-world factory systems, or virtual systems such as a digital psychological test or expert system, require relatively less extensive testing due to their reliance on the assumptions of a closed world.

### 2.4. Rise of the non-professional users

As chips get smaller and gain more capacity, smart and automated technology is becoming more widely available for use by non-professional users. These users have often not been trained in how to use or operate the system and often do not have a detailed technical understanding of how the automation works and the limitations on its successful operation. The last trend that we observe is then that there has been an increase in research on automation for use outside of professional settings. For example, the availability of smart phones and other smart devices that are connected to the internet and allow users to interact with automated systems and processes. Some examples that are covered in IJHCS include electronic shopping (e.g., Hassanein and Head, 2007), robots as social companions (e.g., Leite et al., 2013), and control of semi-automated vehicles (e.g., Rajaonah et al., 2008; Seppelt and Lee, 2007).

While many of the topics that apply to professional (skilled) users of automated systems also apply to non-professional users, there are some additional considerations that come into play for research on how non-professional users interact with automated systems. For example, for non-professional users one cannot rely on extensive training and experience with the technology, and the technology might be used in a wider set of context than that which can be predicted by the profession. Study of use by non-professional users is therefore an emerging setting, discussed in more detail below that requires the full breadth of HCI expertise. Moreover, the use by non-professional users requires further consideration of more ethical topics such as human attitudes towards and acceptance of autonomous systems (Złotowski et al., 2017) and how to handle security and hacking (Chen et al., 2018; Ferreira and Teles, 2019).

### 2.5. Summary of human-automation interaction research to date

In summary, our analysis of publications in IJHCS on the topic of human-automation interaction shows that research has expanded beyond the use of automation in dedicated domains such as factory assembly lines and automated psychological tests. In particular, there are distinct research lines that investigate the use of automation in time-sensitive or safety-critical settings, through embodied situated agents, and by non-professional users.

Fig. 1 provides a Venn diagram with examples of automated systems for each of these research lines. The Venn diagram also makes explicit how these different areas fit together. Specifically, it identifies that there are many domains and settings in which two or more of these research lines come together. A prime example is the automated car, which involves automation in the form of an embodied, situated agent, which is used by non-professional users in a time-sensitive, safety-critical context.

For embodied, situated systems some form of automation (or autonomy) is almost always required (although by definition, humans can also be considered embodied situated agents, Pfeifer and Scheier, 2001). Hence in our Venn Diagram of Fig. 1, embodied, situated agents are represented as a subset of the larger automation category. Moreover, whether something is considered embodied and situated might at times be open to interpretation. For example, we opted that a power plant monitoring system is not labeled as embodied and situated, even though such systems can sense and act to maintain a balance in the power plant's processes (e.g., increase or decrease cooling). Our motivation for not including it as a fully embodied, situated agent was that—from our understanding—these systems tend to rely on if-then rules and are less open to dynamic situations that our other examples (e.g., cars and military drones) face.

## 3. Future of human-automation interaction: evergreen themes

We now turn our attention to the future of human-automation interaction research, by describing themes that are important for future work. We start by describing three themes that are “evergreens”: themes that were also covered in the past, but that continue to be important areas for research. In particular, these themes require further expansion due to the breadth of domains and users that are involved in automated settings. After discussing these evergreen topics, we go on to discuss five new topics in human-automation interaction that we expect to increase in importance over the coming years.

### 3.1. Function and task allocation between humans and machines

The first theme that has had persistent attention in IJHCS research on automation is the distribution or allocation of tasks between humans and automated systems (e.g., Dearden et al., 2000; Hollnagel and Bye, 2000; Press, 1971; Sheridan, 2000; de Vries et al., 2003; Milewski and Lewis, 1997). A simple, naive understanding of the introduction of automation might be that automated systems take over the execution of tasks from humans, and thereby simply ‘reduce’ the amount of work or attention that humans need to dedicate to that task. A colloquial understanding is for example that people are better at some tasks (e.g., to exercise judgment) and machines are better at other tasks (e.g., to perform repetitive routine tasks; Fitts, 1951). However, as analyzed in detail by Sheridan (2000), achieving such allocation in practice is a hard problem, as researchers differ in what they set as appropriate criteria for the function allocation.

In line with this view, it is important to consider the so-called “irony of automation” (Bainbridge, 1983), which states that introduction of automation can radically change how people perceive or act in a specific context. People do not merely reduce what they work on when (part of) a task is automated, but use different strategies for working on that task altogether. For example, one intention of semi-automated vehicles is that the human driver is responsible for fewer basic control-monitoring tasks (e.g., steering, pressing the gas), and can therefore switch his or her attention to monitoring the traffic environment and the vehicle. However, a meta-review of research on driving assistance systems suggests that the introduction of automation increases the likelihood that drivers perform non-driving related tasks, which reduces their situational awareness and response time to alerts (de Winter et al., 2014).

Although the problem of function allocation, and related themes, such as the irony of automation, have been known for decades, the associated research questions gain new urgency now that automation is being used by non-professional users in time-sensitive and safety-critical contexts. An underestimation of user interaction in these domains can lead to incidents, and non-professional users might lack the training and experience to cope with system failures. Moreover, they might underestimate risks or misplace their trust in the system. For example,

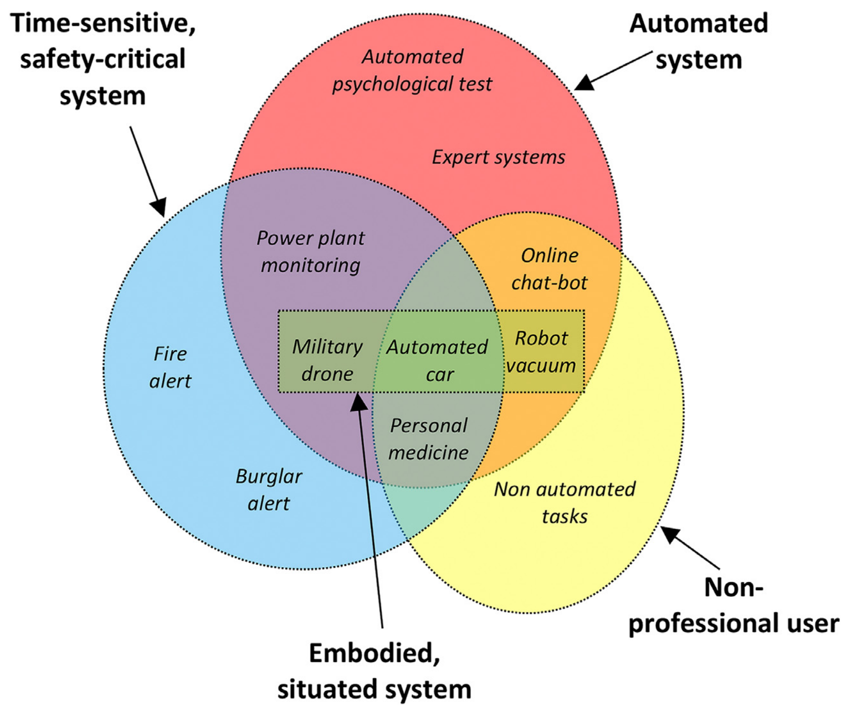


**Fig. 1.** Venn Diagram of current types of human-automation interaction research (not to scale). Automated systems are developed for use by non-professional users, in time-sensitive or safety-critical systems. Embodied situated systems are a subset of automated systems that have seen a rise since the early 2000s.

in the first deadly incident with a Tesla model S (a partially automated vehicle), the human driver had a prolonged period of visual distraction shortly before the crash (Habib, 2017). Although the cause of this distraction is unknown, misplaced trust in the automation might have been a factor.

Automation might also change how, when, and where tasks are performed. For example, if cars become more automated, will they turn into mobile offices (Chuang et al., 2018), or areas of fun and play (Kun et al., 2016)? That is, automation might be a radical disruptive innovation that changes more than just the task itself.

### *3.2. Trust, incorrect use, and confusion*

The second major theme of human-automation interaction to have received persistent attention in IJHCS over the years is how to promote appropriate levels of trust in automation (Dzindolet et al., 2003; Lee and Moray, 1994), how to avoid incorrect use and (human) errors (e.g., Bahner et al., 2008; Skitka et al., 1999, 2000), and how to avoid confusion.

Parasuraman and Riley (1997) introduced four distinct types of use of automation that can impact a user's trust in a system. Initial *use* might already depend on trust, but on top of that users and other stakeholders of automation might *misuse* the automation (i.e., show overreliance, or too much trust), *disuse* it (i.e., under rely on the automation and distrust it, for example due to false alarms), or *abuse* it (i.e., introducing the automation without considering all the consequences of it, in line with the irony of automation, Bainbridge, 1983). These four forms of use, and their impact on trust are still relevant today. They are particularly relevant now that non-professional users are using automation in more settings. As they lack the training and experience of professional users, they might bring in incorrect expectations of the capabilities of the automated system, resulting in misuse or disuse.

How a user uses automation, and how they perceive trust can also be looked at more dynamically, based on a user's understanding of the system's mode of operation over time. The mode, or state, of an automated system determines its response to user input and to changes in the overall context of the system. For example, in automated vehicles, cruise control and adaptive cruise control can be two automation modes. When human drivers or operators engage adaptive cruise control, their vehicle will attempt to maintain a given speed, but will slow down if there is slower traffic ahead; in contrast the same vehicle with (non-adaptive) cruise control will not slow down for slower vehicles ahead. The human operator needs to keep track of mode changes, and also remember how the system will react to user input and context changes in the current mode. Mode confusion (mode error) occurs when the human operator is confused about the current mode of the system, or cannot remember how the system will react in the current mode (Sarter and Woods, 1992).

Mode confusion is highly consequential for safety-critical systems, such as road vehicles, power plants, airplanes, robotic wheelchairs, and flight control systems. In the above example, if the driver mistakenly believes that the vehicle is in the adaptive cruise control mode, when it is actually in (non-adaptive) cruise control (i.e., a form of misuse of automation in Parasuraman and Riley's terms), the result can be a crash. Janssen et al. (2019) discuss this issue in the driving domain by introducing a probabilistic (Hidden Markov Model) framework that relates driver beliefs of the system's mode to actual system modes. Such frameworks make explicit in what system states mode confusion might occur, and can aid in the (re-) design of safety-critical systems.

Mode confusion can also happen in other contexts. Vicente et al. (2001) point out that power plants are highly complex systems, which means that some part of the plant will always be under repair or in a state of being modified. This effectively changes the mode, or state, of the plant, and requires operators to act accordingly. Mode confusion might result in a misinterpretation of alarms: depending on the mode of the power plant, an alarm might indicate an actual problem or an expected state of operation.

In the coming years, human interactions with automation will continue to be subject to mode confusion. The reason is twofold. First, automation is not the same as autonomy: our automated systems will be very good at what they do, but in some difficult cases, or in legally mandated situations, they will require human intervention. Second, automated systems will continue to be applied in a variety of complex situations—after all, that is where they are the most useful. However, use in complex situations will result in multiple modes of operation (Sarter and Woods, 1995). Researchers need to focus on creating models of mode confusion for different application areas, (e.g. Janssen et al., 2019). Such models can then be used in the design and evaluation of systems that reduce the frequency, and the consequences, of these errors.

### 3.3. Focus, divided attention, and attention management

A third theme that has had persistent attention in IJHCS research on automation is creating appropriate workload levels for the human interacting with automation so as to avoid under- and overload (Van Gigh, 1971; Rajaonah et al., 2008). Taking a broader perspective, one can say there is a need to understand focus, divided attention, and attention management.

As automation continues to improve, automated tasks might require less human attention and intervention. This allows humans to focus on other activities, such as (other) work and play. At the same time, researchers expect that humans will continue to play a role in automated systems such as cars, even under higher levels of automation (e.g., Janssen et al., 2019; Lee et al., 2017; Noy et al., 2018; Stone et al., 2016). For example, occasional human aid might be needed if the automated system encounters an off-nominal scenario. In such a case, humans need to revert their attention to the automated task, even though they might feel that their preceding task was more urgent to them. These situations require a detailed understanding of multitasking and interleaving processes (see also special issue in IJHCS, Janssen et al., 2015), and a new view on attention management.

Focusing on automated vehicles, a large body of research has investigated the effectiveness of providing last-minute alerts to warn drivers about situations where human assistance is needed. However, in such automated circumstances, people's susceptibility to alerts is reduced (Van der Heiden et al., 2018; Lahmer et al., 2018; Scheer et al., 2018). Moreover, even if an alert is processed, mode confusion might limit the human driver's understanding of their role and limit their ability to take the right action (Janssen et al., 2019). Novel perspectives on attention management might be needed to minimize these dangers. For example, in our own work we have investigated the use of earlier warnings (pre-alerts) to warn drivers before their action is critical (Van der Heiden et al., 2017; see also Borojeni et al., 2018). Beyond simply providing warnings, more research is needed into how the human and the machine can be *partners* in a task, instead of one taking over the task of the other and only warning in case of emergency. The success of such systems will rely both on the system's ability to assess (e.g., model and predict) the human state and understanding, and also on the human's ability to understand the system's functioning.

## 4. Future of human-automation interaction: emerging themes

To close, we discuss five themes that are emerging as important topics in automation research, and which we expect to increase in importance over the years to come.

### 4.1. Interdisciplinary studies to cover breadth and depth of domains and users

Our review of the IJHCS literature has shown that over the past five decades, research on human-automation interaction has broadened out into different areas. We expect that automated systems will continue to broaden out into new domains as the principles and methods behind automated technologies aimed at professional users start to penetrate the broader consumer market aimed at non-professional users. For example, automated features from commercial airplanes might make it over to non-commercial airplanes that are used by trained, but less experienced pilots.

At the same time, even though technology branches out, in a sense automated technology is often still specialized and limited, and its accuracy can be improved. In the home environment there are dedicated machines for vacuuming, lawn mowing, or playing music, but few devices that combine such tasks. Personal virtual assistants like Amazon's Alexa, Apple's Siri, or Google Assistant can aid in many tasks, but have limited capabilities (e.g., Cohen et al., 2016; Cowan et al., 2017). On the road, automated cars can tackle ever more complex and demanding situations, but still have exceptions where human assistance is needed. In other words, there are opportunities for improvements in both the "depth" (i.e. improving performance on specific tasks) and the "breadth" (i.e., how many tasks and contexts they can handle) of studies on automated systems.

As part of the branching out, automated systems will be used more frequently by non-professional users and with this comes a set of important questions about human-automation interaction. For example, how are users trained to work with automated safety-critical devices? How are their skills on a task retained if it is not put to use frequently (see also Casner et al., 2014)? How are different cultures, and different norms, customs, and conventions facilitated? Will the adoption and use of automated systems benefit a variety of user groups (e.g., automated vehicles hold the potential for improved mobility for people who cannot drive or do not have access to their own vehicle)?

### 4.2. Regulation and explainability

The regulatory landscape for automation depends heavily on the application area. Thus, regulation is well-developed for established fields, such as for relatively simple medical devices. However, new interconnected medical devices present a challenge for regulation (Sokolsky et al., 2011). Even more so, medical robotics, where automation can take on various forms, presents a significant challenge for regulators—in fact, autonomous robots will not only be medical devices but also entities that practice medicine, and it is not yet clear who would be in charge of regulating them (Yang et al., 2017). Similarly, regulation is still under development for cars, where automation is only now making significant advances (Inners and Kun, 2017).

A large push on automation research comes from European legislation on "explainability". In the context of recent data protection laws, European laws now require that decisions that are made for humans by automated systems are explainable to the humans (European Union, 2016, 2018; see also Goodman and Flaxman, 2017). Automated system and (machine learning) algorithms make many decisions, but the reasons for these decisions might be opaque to the end user (Burrell, 2016). Moreover, the (decision) models that the algorithms create to inform their actions necessarily abstract away from some details in the world. Such abstraction can result in 'traps' (Selbst et al., 2018) such as an inability to take all of the relevant features into account in decision making (as some were left out in the abstraction) or to transfer learned behavior to new settings (where other features are perhaps more important).

Explainability is not always straightforward for embodied, situated automated systems such as automated cars, as these systems make many decisions over time. For example, at any given time there is an explicit or implicit decision to accelerate or decelerate, and whether to make a steering adjustment (i.e., Michon's control level; Michon, 1985). Should cars be able to explain these decisions continuously? And should this be done in real-time? Or should only more strategic decisions (Michon, 1985) such as why particular routes were chosen be explainable? Or is only hindsight explanation needed surrounding (near-) accidents? Although ideally a system should be able to make multiple explanations, whether they do this can impact a user's attention, and might also have impact on system performance (i.e., when dedicating capacity to the storing of decisions). From a human-computer interaction perspective, explainability of automated systems should at least be present to avoid mode confusion (Janssen et al., 2019) and to avoid alert fatigue and the so-called "cry-wolf effect" (Breznitz, 1983; Sorkin, 1989).

Like humans, automated machines are not always "perfect". The algorithms behind automated systems often get trained on data, and the resulting decision systems might be limited by the data ("Garbage in, garbage out"). Specifically, through the training set, the algorithms might pick up on biases or inequalities that exist in society, which can have consequences for the end users. For example, if a gender classification algorithm is trained to classify people based on their physical features, it might overlook that biological sex and self-identified gender labels might not align, and the resulting misgendering might have negative impacts on mental health (Hamidi et al., 2018).

Humans might be able to help learning systems to overcome their biases. For example, in recently proposed guidelines for human-AI interaction, five of the eighteen guidelines focus on ways to help users correct the mistakes of an AI system (Amershi et al., 2019). However, it is an open question how to design such systems in practice, in particular as there might be a disconnect between the low-level features that a system needs to adjust to improve, and the high-level concepts that a user (incorrectly) thinks they need to adjust (e.g., Kittley-Davies et al., 2019).

From a legislative perspective, an important question is then also who is to blame when an accident or incident occurs involving an automated system in a safety-critical setting. An initial thought might be to think locally, with the human operator or the producer, programmer, or seller of the technology. However, the introduction of automation is sometimes motivated by a narrative to reduce the frequency or probability of accidents and incidents. Approaching these from a probabilistic viewpoint raises the question of what is an acceptable probability of risk, and how this risk is spread over the population. The consideration of risk at the population level, then turns the question of "who is to blame" into a question that is probably larger than one individual.

### 4.3. Ethical and social dilemmas

As automated machines achieve more functionality, various ethical and social dilemmas become more urgent and prominent. Our overview of the history of IJHCS already touched on one such issue: are increasingly autonomous systems socially accepted as equals (Złotowski et al., 2017)?

Another ethical and social consideration is that of the future of work and job security. A model by Frey and Osborne (2017) predicts that low-skill and low-wage jobs, such as in transportation, logistics, and office work, in particularly are likely to be replaced by automation. Frey and Osborne predict that this will require a shift in skillsets by human workers to tasks that require creativity or social skills. From our perspective, it is unclear whether this prediction will hold, as our literature review of IJHCS articles indicates that research is already investigating topics such as emotion classification and social interaction between humans and robots (e.g., Brave et al., 2005; Hassanein and Head, 2007; Kapoor et al., 2007; Leite et al., 2013). Therefore, we expect that in the years to come there will be more progress on (partial) automation of creative tasks and social interaction settings than anticipated in the report by Frey and Osborne. If this happens, the ethical and social question of job security will be plainly evident.

Moreover, automation might not increase at a steady, linear pace. For example, Harari (2018) predicts that the pace of improvements in automation might also accelerate as time goes on, thereby making it ever harder for people to catch up with the increasing changes in automation and to adapt their skillset. How are humans then equipped for these societal changes? How do we make sure that we create devices that are there for human users? But also, how can technology help to achieve a world that provides opportunity for all, and not just for a fortunate minority?

Another ethical consideration is what decisions automated systems should take in complex life-or-death situations that are imminent in safety-critical scenarios. Survey research shows that humans would like automated machines to make morally just decisions in principle, yet they also want the system to deviate from this moral path if a moral action would require sacrificing their own life or that of their family members (Bonnefon et al., 2016). Moreover, the survey research shows that there are individual and cultural differences in what is considered morally just (Awad et al., 2018). Given that humans cannot agree on moral conflicts, a lot more research is needed to guide the regulation of automated systems. For example, the Ethics Commission on Automated and Connected Driving, which was appointed by the German government, has developed a set of twenty ethical rules related to the design, deployment, legal issues, and use of automated vehicles (Ethics Commission, 2017).

Taken together, the full set of social and ethical considerations also poses a fundamental question: whether to automate at all or not? In most safety-critical scenarios where automation is introduced, such as automated driving, the intention is that introduction of automation or automated support can save lives and reduce incidents. However, the new technology can also introduce new problems and incidents. A moral judgment is needed whether the benefits weigh up against the challenges. Although the inclination of some researchers might be to minimize *new* incidents, this might overlook the benefits of automation (see also de Winter, 2019).

### 4.4. Continued and improved human and humane experiences

Implicit in the previously discussed trends is the need to consider human experience. With automation improving, how can we continue to maintain a fair and humane interaction (see also section on ethics)? Which aspects of tasks do we automate, and which tasks do we leave to the human? In line with the historical trend of automated testing (e.g., Elithorn et al., 1982; Thompson and Wilson, 1982) and expert systems (Motta, 2013; Gaines, 2013; Breuker, 2013), we might expect more software tasks to become automated in the coming few years. But which parts are automated? How is creativity and expertise embedded correctly? If creativity is essential for human contributions to an automated task, how do we ensure that humans can contribute this, and how do we know when and where it is needed? Or, if humans would like to focus on other aspects of a task, apart from creativity, how do we continue to allow them to do so? For example, in a world where automated vehicles have penetrated the market, will we allow occasional human driving "just for fun"? How can this be done in a world where other cars might rely on the predictability of non-human actions to maintain a stable driving trajectory? If we do not allow humans to contribute to such tasks and activities, how do we allow a humane experience in other ways? The answers to these questions are not yet clear, but needed.

### 4.5. Radical changes to human-automation interaction

As we look into the future, technological advances in human-machine interaction, automation, artificial intelligence, and related disciplines are likely to usher in dramatic change in how we live with computing devices. Although such radical shifts are hard to predict accurately, some suggestions and trends are noticeable.

One such change is imagined by Yuval Noah Harari in his book "21 lessons for the 21st century" (Harari, 2018)—he envisions a world in which AI will become better than we are at many tasks. If this happens, then one question for human-automation design will be how human users can best use such super-smart AI. Will the humans enjoy the interactions and engage in them? Will they engage with AI while having the appropriate level of trust, taking into account both the benefits and the potential costs of the interactions? Or will they act like the humans in Asimov's (1954) novel "Caves of Steel," where the people of the Earth of about 1000 years in the future fear and reject robots, and the comforts that robots can provide humanity?

Another dramatic change is envisioned by the futurist Mark Pesce—he expects that we will be able to associate digital data with physical objects and view this data through augmented reality glasses (Pesce, 2019). Pesce expects that this will lead to the emergence of what he calls 'supertools': tools that can allow us to interact with computing objects, and thus with the automation around us, while having at our disposal vast amounts of data about all aspects of the work of automation. One significant question for human-computer interaction design in this case is how to allow users to interact with this vast amount of data. Simply put, there will be too much data available for users to be able to handle it all, which means that human-computer interaction design will need to create focused views of the data.

Turning to art again, and specifically the science fiction of Asimov: imagine what it might be like to interact with automation if our interface technologies can go beyond showing us information with augmented reality! What if the interfaces could make us feel like the machine is an extension of our body? This is what it feels to operate an advanced starship in Asimov's (1982) "Foundation's Edge"—the effort required to accomplish something is about as much as to think about the goal. Perhaps Asimov overestimated the probability that machines will eventually be able to literally read our minds. But, we can still expect that our minds and the machine automation will not always be separated by keyboards, screens, and brittle speech interfaces. How will radically more capable interfaces affect how we can control automation, and just as importantly, how we perceive automation and its place in our lives?

As we contemplate the inevitable radical changes in human-automation interaction, it is important to keep asking questions. What are the economic and societal forces that are driving the changes? How will new technologies shape what is possible for these interactions? And what are the economic and broad societal implications of these dramatic changes? The answers to these questions will be found through interdisciplinary work that incorporates a clear understanding of human-automation interaction, and leverages it effectively.

Many previous eras of human development have included radical change in technology, but we expect the change to be faster than it had been in the past. Where will this change lead us? For all of the themes we mentioned in this document, except for this last one, we have reasonably clear plans for how to move forward. For some of them, our horizon extends relatively far, for others not that far.

In sum, human-automation interaction research has been an area of exciting and impactful work for many decades. The readers of IJHCS, and more broadly the scientific community, should expect this trend to accelerate in the coming years.

## Declarations of interest

None.

## References


Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S.T., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., 2019. Guidelines for human-AI interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY Paper 3.

Asimov, Isaac, 1954. The Caves of Steel. Doubleday, New York, NY.

Asimov, Isaac, 1982. Foundation's Edge. Doubleday, New York, NY.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F., Rawan, I., 2018. The moral machine experiment. Nature 563 (7729), 59–64.

Bahner, J.E., Hüper, A.D., Manzey, D., 2008. Misuse of automated decision aids: complacency, automation bias and the impact of training experience. Int. J. Hum. Comput. Stud. 66 (9), 688–699.

Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A., Nass, C., John, O., 2008. Real-time classification of evoked emotions using facial feature tracking and physiological responses. Int. J. Hum. Comput. Stud. 66 (5), 303–317.

Bainbridge, L., 1983. Ironies of automation. Automatica 19, 775–780.

Belkin, N.J., Brooks, H.M., Daniels, P.J., 1987. Knowledge elicitation using discourse analysis. Int. J. Man Mach. Stud. 27 (2), 127–144.

Bonnefon, J.F., Shariff, A., Rahwan, I., 2016. The social dilemma of autonomous vehicles. Science 52 (6293), 1573–1576.

Borojeni, S.S., Weber, L., Heuten, W., Boll, S., 2018. From reading to driving: priming mobile users for take-over situations in highly automated driving. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (article 14). ACM Press, New York, NY.

Brave, S., Nass, C., Hutchinson, K., 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int. J. Hum. Comput. Stud. 62 (2), 161–178.

Breuker, J., 2013. A cognitive science perspective on knowledge acquisition. Int. J. Hum. Comput. Stud. 71 (2), 177–183.

Breznitz, S., 1983. Crywolf: The Psychology of False Alarms. Lawrence Erlbaum, Hillsdale, NJ.

Burrell, J., 2016. How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data Soc. 3 (1), 2053951715622512.

Carroll, J.M. , 2013. Human computer interaction-brief intro. In: The Encyclopedia of Human-Computer Interaction, 2nd Ed. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro.

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

Chen, J., Mishler, S., Hu, B., Li, N., Proctor, R.W., 2018. The description-experience gap in the effect of warning reliability on user trust and performance in a phishing detection context. Int. J. Hum. Comput. Stud. 119, 35–47.

Chien, S.Y., Lin, Y.L., Lee, P.J., Han, S., Lewis, M., Sycara, K., 2018. Attention allocation for human multi-robot control: cognitive analysis based on behavior data and hidden states. Int. J. Hum. Comput. Stud. 117, 30–44.

Chuang, L.L., Donker, S.F., Kun, A.L., Janssen, C.P., 2018. Workshop on the mobile office. In: Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM Press, New York, NY, pp. 10–16.

Cohen, P., Cheyer, A., Horvitz, E., El Kaliouby, R., Whittaker, S., 2016. On the future of personal assistants. In: Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM Press, New York, NY, pp. 1032–1037.

Cowan, B.R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Early, D., Bandeira, N., 2017. What can I help you with?: infrequent users' experiences of intelligent personal assistants. In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM Press, New York, NY, pp. 43.

de Vries, P., Midden, C., Bouwhuis, D., 2003. The effects of errors on system trust, self-confidence, and the allocation of control in route planning. Int. J. Hum. Comput. Stud. 58 (6), 719–735.

de Winter, J.C.F., 2019. Pitfalls of automation: a faulty narrative? Commentary on Hancock (2019) some pitfalls in the promises of automated and autonomous vehicles. Ergonomics 1–4.

de Winter, J.C.F., Happee, R., Martens, M.H., Stanton, N.A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. Transp. Res. Part F 27, 196–217.

Dearden, A., Harrison, M., Wright, P., 2000. Allocation of function: scenarios, context and the economics of effort. Int. J. Hum. Comput. Stud. 52 (2), 289–318.

Diederich, J., Ruhmann, I., May, M., 1987. KRITON: a knowledge-acquisition tool for expert systems. Int. J. Man Mach. Stud. 26 (1), 29–40.

Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P., 2003. The role of trust in automation reliance. Int. J. Hum. Comput. Stud. 58 (6), 697–718.

Elithorn, A., Mornington, S., Stavrou, A., 1982. Automated psychological testing: some principles and practice. Int. J. Man Mach. Stud. 17 (3), 247–263.

Ethics Commission (2017) Automated and connected driving. Federal ministry of transport and digital infrastructure, Germany. Available online at: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.html?nn = 187598.

European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). http://data.europa.eu/eli/reg/2016/679/oj.

Feigenbaum, E.A., The art of artificial intelligence: 1. Themes and case studies of knowledge engineering. No. STAN-CS-77-621. Stanford Univ CA Dept of Computer Science, 1977. Available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/a046289.pdf.

European Union (2018) Corrigendum to Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). http://data.europa.eu/eli/reg/2016/679/corrigendum/2018-05-23/oj.

Ferreira, A., Teles, S., 2019. Persuasion: how phishing emails can influence users and bypass security measures. Int. J. Hum. Comput. Stud. 125, 19–31.

Fitts, P.M., 1951. Human Engineering for an Effective Air Navigation and Traffic Control System. National Research Council, Washington, DC.

Frey, C.B., Osborne, M.A., 2017. The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Change 114, 254–280.

Gaines, B.R., 2013. Knowledge acquisition: past, present and future. Int. J. Hum. Comput. Stud. 71 (2), 135–156.

Goodman, B., Flaxman, S., 2017. European Union regulations on algorithmic decision-making and a "right to explanation". AI Mag. 38 (3), 50–57.

Gould, S.J., Cox, A.L., Brumby, D.P., 2018. Influencing and measuring behaviour in crowdsourced activities. In: Filimowicz, M., Tzankova, V. (Eds.), New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies. Springer, pp. 103–130.

Groover, M.P., 2018. Automation. Encyclopædia Britannica. Encyclopædia Britannica,

inc. Accessed online on 26 January 2019, at. https://www.britannica.com/technology/automation.
Gruber, T.R., Cohen, P.R., 1987. Design for acquisition: principles of knowledge-system design to facilitate knowledge acquisition. Int. J. Man Mach. Stud. 26 (2), 143–159.
Habib, K., 2017. Automatic Vehicle Control Systems. National Highway Traffic Safety Administration, Department of Transportation, pp. 1–13. Technical Report PE 16-007pages. Available online at: https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF.
Hamidi, F., Scheuerman, M.K., Branham, S.M., 2018. Gender recognition or gender reductionism?: the social implications of embedded gender recognition systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (paper 8). ACM Press, New York, NY.
Harari, Y.N., 2018. 21 Lessons for the 21st Century. Random House.
Hassanein, K., Head, M., 2007. Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. Int. J. Hum. Comput. Stud. 65 (8), 689–708.
Hollnagel, E., Bye, A., 2000. Principles for modelling function allocation. Int. J. Hum. Comput. Stud. 52 (2), 253–265.
Inners, M, Kun, A, 2017. Beyond liability: legal issues of human-machine interaction for automated vehicles. In: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM Press, New York, NY, pp. 245–253.
Janssen, C.P., Boyle, L., Kun, A.L., Ju, W., Chuang, L.L., 2019. A hidden markov framework to capture human–machine interaction in automated vehicles. Int. J. Hum. Comput. Interact. https://doi.org/10.1080/10447318.2018.1561789.
Janssen, C.P., Gould, S.J.J., Li, S.Y.W., Brumby, D.R., Cox, A.L., 2015. Integrating knowledge of multitasking and interruptions across different perspectives and research methods. Int. J. Hum. Comput. Stud. 79, 1–5.
Kaber, D.B., Wright, M.C., Sheik-Nainar, M.A., 2006. Investigation of multi-modal interface features for adaptive automation of a human–robot system. Int. J. Hum. Comput. Stud. 64 (6), 527–540.
Kapoor, A., Burleson, W., Picard, R.W., 2007. Automatic prediction of frustration. Int. J. Hum. Comput. Stud. 65 (8), 724–736.
Kittley-Davies, J., Alqaraawi, A., Yang, R., Costanza, E., Rogers, A., Stein, S., 2019. Evaluating the effect of feedback from different computer vision processing stages: a comparative lab study. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Proceedings. ACM Press, New York, NY Paper 43.
Kun, A.L., Boll, S., Schmidt, A., 2016. Shifting gears: user interfaces in the age of autonomous vehicles. IEEE Pervasive Comput. 15, 32–38.
Lahmer, M., Glatz, C., Seibold, V.C., Chuang, L.L., 2018. Looming auditory collision warnings for semi-automated driving: an ERP study. In: Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM Press, New York, NY, pp. 310–319.
Lee, J.D., Moray, N., 1994. Trust, self-confidence, and operators' adaptation to automation. Int. J. Hum. Comput. Stud. 40 (1), 153–184.
Lee, J.D., Wickens, C.D., Liu, Y., Boyle, L.N., 2017. Human-automation interaction. In: Lee, Wickens, Liu, Boyle (Eds.), Designing for People: An Introduction to Human Factors Engineering. CreateSpace, Charleston, SC.
Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., Paiva, A., 2013. The influence of empathy in human–robot relations. Int. J. Hum. Comput. Stud. 71 (3), 250–260.
Madrigal, A. (2017). Inside Waymo's secret world for training self-driving cars. The Atlantic (August 23, 2017). Available online at:https://www.theatlantic.com/technology/archive/2017/08/inside-waymos-secret-testing-and-simulation-facilities/537648/.
Michon, J.A., 1985. A critical view of driver behavior models: what do we know, what should we do? Human Behavior and Traffic Safety. Springer, Boston, MA, pp. 485–524.
Milewski, A.E., Lewis, S.H., 1997. Delegating to software agents. Int. J. Hum. Comput. Stud. 46 (4), 485–500.
Motta, E., 2013. 25 years of knowledge acquisition. Int. J. Hum. Comput. Stud. 71 (2), 131–134.
Noy, I.Y., Shinar, D., Horrey, W.J., 2018. Automated driving: safety blind spots. Saf. Sci. 102, 68–78.
Parasuraman, R., Riley, V., 1997. Humans and automation: use, misuse, disuse, abuse. Hum. Factors 39 (2), 230–253.
Partala, T., Surakka, V., 2003. Pupil size variation as an indication of affective processing. Int. J. Hum. Comput. Stud. 59 (1–2), 185–198.
Pesce, M. (2019). Supertools. Retrieved on 26 April 2019 fromhttps://youtu.be/ftYgdpnfc-g.
Pfeifer, R., Scheier, C., 2001. Understanding Intelligence. MIT press, Cambridge, MA.
Press, L., 1971. Toward balanced man-machine systems. Int. J. Man Mach. Stud. 3 (1), 61–73.
Rajaonah, B., Tricot, N., Anceaux, F., Millot, P., 2008. The role of intervening variables in driver—ACC cooperation. Int. J. Hum. Comput. Stud. 66 (3), 185–197.
Russell, S., Norvig, P., 2009. Artificial Intelligence: A Modern Approach, 3rd ed. Prentice Hall Press, Upper Saddle River, NJ, USA.
Sakamoto, D., Kanda, T., Ono, T., Kamashima, M., Imai, M., Ishiguro, H., 2005. Cooperative embodied communication emerged by interactive humanoid robots. Int. J. Hum. Comput. Stud. 62 (2), 247–265.
Sarter, N.B., Woods, D.D., 1992. Mode error in supervisory control of automated systems. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 36. pp. 26–29 Sage CA: Los Angeles, CA: SAGE Publications, 1992.
Sarter, N.B., Woods, D.D., 1995. How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum. Factors 37 (1), 5–19.
Scheer, M., Bülthoff, H.H., Chuang, L.L., 2018. Auditory task irrelevance: a basis for inattentional deafness. Hum. Factors 60 (3), 428–440.
Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., Vertesi, J., 2018. Fairness and abstraction in sociotechnical systems. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. ACM Press, New York, NY, pp. 59–68.
Seppelt, B.D., Lee, J.D., 2007. Making adaptive cruise control (ACC) limits visible. Int. J. Hum. Comput. Stud. 65 (3), 192–205.
Sheridan, T.B., 2000. Function allocation: algorithm, alchemy or apostasy? Int. J. Hum. Comput. Stud. 52 (2), 203–216.
Singh, I.L., Molloy, R., Parasuraman, R., 1997. Automation-induced monitoring inefficiency: role of display location. Int. J. Hum. Comput. Stud. 46 (1), 17–30.
Skitka, L.J., Mosier, K., Burdick, M.D., 2000. Accountability and automation bias. Int. J. Hum. Comput. Stud. 52 (4), 701–717.
Skitka, L.J., Mosier, K.L., Burdick, M., 1999. Does automation bias decision-making? Int. J. Hum. Comput. Stud. 51 (5), 991–1006.
Sokolsky, O., Lee, I., Heimdahl, M., 2011. Challenges in the regulatory approval of medical cyber-physical systems. In: Proceedings of the Ninth ACM International Conference on Embedded Software, pp. 227–232.
Sorkin, R.D., 1989. Why are people turning off our alarms. Hum. Factors Bull. 32 (4), 3–4.
Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., et al., 2016. Artificial Intelligence and Life in 2030. Stanford University, Stanford, CA.
Thompson, J.A., Wilson, S.L., 1982. Automated psychological testing. Int. J. Man Mach. Stud. 17 (3), 279–289.
van der Heiden, R.M., Iqbal, S.T., Janssen, C.P., 2017. Priming drivers before handover in semi-autonomous cars. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM Press, New York, NY, pp. 392–404.
van der Heiden, R.M., Janssen, C.P., Donker, S.F., Hardeman, L.E., Mans, K., Kenemans, J.L., 2018. Susceptibility to audio signals during autonomous driving. PloS One 13 (8), e0201963.
van Gigch, J.P., 1971. Changes in the mental content of work exemplified by lumber sorting operationst. Int. J. Man Mach. Stud. 3 (1), 13–29.
Vicente, Kim J., Roth, E.M., Mumaw, R.J., 2001. How do operators monitor a complex, dynamic work domain? The impact of control room technology. Int. J. Hum. Comput. Stud. 54 (6), 831–856.
Yang, G.Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P.E., et al., 2017. Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy. Sci. Robot. 2 (4), 8638.
Zeng, D., Sycara, K., 1998. Bayesian learning in negotiation. Int. J. Hum. Comput. Stud. 48 (1), 125–141.
Złotowski, J., Yogeeswaran, K., Bartneck, C., 2017. Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. In. J. Hum. Comput. Stud. 100, 48–54.

**Christian P. Janssen** is an assistant professor of experimental psychology at Utrecht University. He received his PhD in human-computer interaction from UCL (2012).

**Stella F. Donker** is an associate professor of experimental psychology at Utrecht University. She received her PhD in medical sciences from the University of Groningen (2002).

**Duncan P. Brumby** is a reader of Human-Computer Interaction at University College London (UCL). He received his PhD in psychology from Cardiff University.

**Andrew L. Kun** is an associate professor of electrical and computer engineering at the University of New Hampshire. He received his PhD in electrical engineering from the University of New Hampshire (1997).