Front. Eng. Manag. 2022, 9(1): 89–103
https://doi.org/10.1007/s42524-021-0182-0

REVIEW ARTICLE

Wei XIONG, Hongmiao FAN, Liang MA, Chen WANG

Challenges of human–machine collaboration in risky decision-making

Abstract The purpose of this paper is to delineate the research challenges of human–machine collaboration in risky decision-making. Technological advances in machine intelligence have enabled a growing number of applications in human–machine collaborative decision-making. Therefore, it is desirable to achieve superior performance by fully leveraging human and machine capabilities. In risky decision-making, a human decision-maker is vulnerable to cognitive biases when judging the possible outcomes of a risky event, whereas a machine decision-maker cannot handle new and dynamic contexts with incomplete information well. We first summarize features of risky decision-making and possible biases of human decision-makers therein. Then, we argue the necessity and urgency of advancing human–machine collaboration in risky decision-making. Afterward, we review the literature on human–machine collaboration in a general decision context, from the perspectives of human–machine organization, relationship, and collaboration. Lastly, we propose challenges of enhancing human–machine communication and teamwork in risky decision-making, followed by future research avenues.

Keywords human–machine collaboration, risky decision-making, human–machine team and interaction, task allocation, human–machine relationship

1 Introduction

The machine in this paper refers to an intelligent system that can make decisions in an autonomous and (partially or fully) independent manner, and the machine's autonomy is realized through artificial intelligence (AI), deep learning, or other algorithms (Rahwan et al., 2019).

With the rapid development of information technology such as AI, deep learning, and big data (Duan et al., 2019), machines have transitioned from mechanization and automation to intelligentization in the past decades. Increased computation power and advanced algorithms enable the machine to catch up with or even surpass human capabilities in various contexts. For example, Alpha Go beats the top human player in a strategy game that used to be exclusive to humans (Silver et al., 2016; 2017). Google's unmanned vehicles are safer and more stable than vehicles driven by humans (Hancock et al., 2020). In medical treatment (Patel et al., 2019; Topol, 2019), intelligence technology has developed unprecedentedly fast to assist medical staff in diagnosing and caring for patients. Given evolutions in the depth and breadth of machine capabilities, machines' ability to undertake increasingly independent and essential tasks in the near future is promising.

However, machines, especially weak AIs, are often developed for specific purposes and trained with limited input data. As a result, they generally perform well only within a pre-defined scope. In addition, the causal relationships between input and output data are often poorly structured; hence, many weak AI machines function as "black boxes". Even though the capability and speed of a machine (or algorithm) in collecting, processing, and analyzing information can easily surpass those of a human decision-maker (Jarrahi, 2018), its flexibility, adaptability, and accountability are lacking. Moreover, autonomous machines still cannot work independently without human supervision in decisions involving high stakes such as human lives (Xu, 2019) due to insufficient reasoning under moral dilemmas.

Received September 9, 2021; accepted November 28, 2021

Wei XIONG, Hongmiao FAN, Liang MA (✉), Chen WANG (✉)
Laboratory of Enhanced Human–Machine Collaborative Decision-Making, Department of Industrial Engineering, Tsinghua University, Beijing 100084, China
E-mails: liangma@tsinghua.edu.cn; chenwang@tsinghua.edu.cn

This study was supported by the National Natural Science Foundation of China (Grant Nos. 71871128, 72171127 and 72192824) and Beijing Social Science Fund (Grant No. 19GLB029).

Currently, integrating human and machine capabilities has demonstrated potential in various applications. For example, the swarm-based technology, i.e., a networked group of radiologists modeled after biological swarms, together with deep-learning technology, was shown to achieve superior diagnostic accuracy than either method alone (Patel et al., 2019). Damacharla et al. (2018) reported that a team of two non-expert chess players and three personal computers outperformed either a group of supercomputers or a group of grandmasters. Holzinger (2016) discussed the efficacy of interactive machine learning (iML), which guides machine computation with human expertise, in solving computationally difficult problems, for example, subspace clustering in protein folding.

Moreover, machines are tools to humans instead of teammates (Phillips et al., 2011). The development of machines has shifted human–machine relationships from mere interaction to cooperation, teaming, and collaboration (Hoc, 2000; Xu, 2019; Haesevoets et al., 2021). In a hybrid system, machine behaviors influence and shape human behaviors and vice versa (Rahwan et al., 2019). Parker and Grote (2019) argued that the fundamental difference between the emerging human–AI relationship and the traditional human–machine relationship posed new questions on work design. Roth et al. (2019) also advocated proactive research on work design and function allocation. Besides, facilitating mutual understanding between humans and machines is a pressing matter. Many recent studies have focused on designing explainable and trustable AIs and building effective and healthy human–AI relationships (Gunning, 2016; DARPA, 2018; Warden et al., 2019).

In this paper, we are particularly interested in summarizing research challenges in human–machine collaborative decision-making under risk. Risky decision-making refers to the problem of making choices without knowing the exact consequences (Bier et al., 1999). Such problems are ubiquitous in economics, technology, policy-making, and daily life, for example, determining a portfolio of investments, allocating medical resources during a pandemic, and deciding which medical treatment to apply. In a typical risky scenario, the decision-maker faces several choices, and each choice involves multiple possible outcomes whose probabilities are knowable. Thus, likelihoods and consequences are two critical dimensions to characterize the outcome of such a decision (Bedford and Cooke, 2001). In a normative perspective, the decision-maker is supposed to act rationally according to expected utility theory (von Neumann and Morgenstern, 1944) based on evaluating the likelihoods and consequences of all possible outcomes. In a descriptive perspective, we often observe cognitive biases in human decisions, and people tend to exercise simple heuristics to reach a solution (Kahneman and Tversky, 1979).

Many open questions arise about establishing human–machine teams for risky decision-making, where the context is generally uncertain, complex, and dynamic. For example, who should be assigned with which tasks, including cognition, judgment, and decision, and under what principles? How can a machine understand human decision-makers' values and behaviors and prescribe both normatively correct and subjectively acceptable solutions? The quest to answer these essential questions motivated us to conduct a literature review and seek potential research directions in this paper.

In particular, we organized this paper as follows. We discussed the necessity and urgency of human–machine collaboration in risky decision-making in Section 2. In Section 3, we reviewed current developments on human–machine collaboration for task allocation (organizational perspective), human–machine relationship (relationship perspective), and human–machine interfaces (interaction perspective). In Section 4, we focused on emerging challenges and potential research avenues concerning human–machine collaboration in risky decision-making. Finally, Section 5 concluded this review paper.

2 Human–machine collaboration in risky decision-making

2.1 Human decision-maker's cognitive limitations in risky decision-making

The outcome of a decision is susceptible to uncertainty (March and Shapira, 1987). Thus, the decision quality also relies on how a decision-maker handles uncertainty, including aleatory uncertainty that is inherent in the random phenomenon underlying a risk event, and epistemic uncertainty that arises from incomplete knowledge about the phenomenon under consideration (Apel et al., 2004). Human decision-makers have been shown to employ simple heuristics and exhibit cognitive biases (see Fig. 1).

When assessing the likelihoods and consequences of a risk event, people are vulnerable to representativeness bias, availability bias, and anchoring effect (Tversky and Kahneman, 1974; Bier, 2004; Kahneman and Frederick, 2002), and their judgments do not always follow the Bayesian rule (Grether, 1992; El-Gamal and Grether, 1995; Griffiths and Tenenbaum, 2006; Charness et al., 2007). In empirical studies, cognitive biases are widely found in the fields of operations management, medical diagnosis, enterprise strategy, and investment (Wickham, 2003; Chen et al., 2007; Croskerry, 2013; Blumenthal-Barby and Krieger, 2015; Tong and Feiler, 2017). Moreover, people fall prey to bounded rationality by not choosing an option with the highest expected utility (von Neumann and Morgenstern, 1944; Wakker, 1989) but following other rules (Edwards, 1962; Luce and Fishburn, 1991; Mearman, 2011) as suggested by prospect theory (Kahneman and Tversky, 1979), regret theory (Bell, 1982), and satisficing

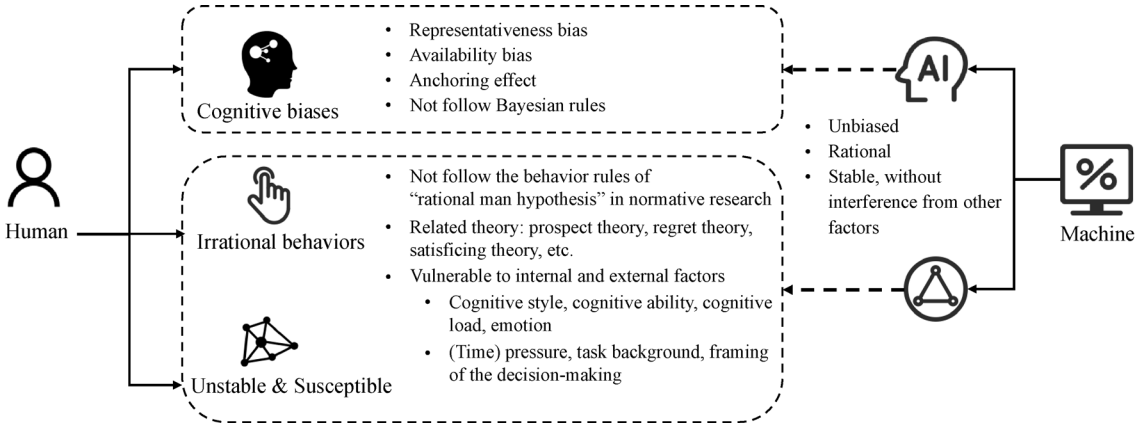

Fig. 1 Human decision-maker's limitations and machine's potential to enhance risky decision-making.

theory (Simon et al., 2004). Additionally, the decision strategies people adopt are dependent not only on the decision-maker's cognitive style (Hunt et al., 1989) and ability (Cokely and Kelley, 2009), but also on the contexts such as cognitive load (Deck and Jahedi, 2015), emotion and pressure (Zinn, 2008; Ordóñez et al., 2015), and framing of the decision task (Payne et al., 1993; Dörner and Wearing, 1995). For example, in public decision-making, consequences will be borne by someone other than the decision-maker (Gregory et al., 1996), leading to strong emotional reactions (Gregory et al., 1996) and moral dilemmas (Tetlock, 2003). In general, cognitive limitations (see Fig. 1) of human decision-makers create barriers to rational and consistent decisions and induce unwanted results.

2.2 Opportunities for human–machine collaboration in risky decision-making

Human–machine collaboration has great potentials for risky decision-making. Machines could be more supportive in gathering information and assessing uncertainties, and conveying key messages to human decision-makers to save cognitive resources. Moreover, human decision-makers could debias their judgments and constrain emotional influences with the assistance of a machine. Empirical evidence has demonstrated the benefits of human–machine collaboration in decisions under risk. Take medical diagnosis as an example. Dawes et al. (1989) compared the performance of doctors' clinical models with the regression-based actuarial model and found that the actuarial model was more accurate than the subjective diagnostic model. Nowadays, the superiority of algorithms remains prevalent in clinical settings (Miller, 2018; Topol, 2019), thanks to their invulnerability to cognitive biases, fatigue, recent experience, and environmental factors (Whelehan et al., 2020). However, doctors consistently outperformed AI in outlier analysis and rapid grasp of new, complex, and rare symptoms (Amann et al., 2020; Lee, 2020). Therefore, the synergy of human and machine capabilities is desirable (Lee, 2020).

In this section, we characterize the opportunities for human–machine collaboration in risky decision-making by the levels of uncertainty involved (see Fig. 2). On the one hand, when the decision task features low uncertainty, the research opportunities are mainly algorithm-centered, which lie in the effective utilization of the computing power of machines (Patel et al., 2019). If programmed in proper ways following decision rules approved by a human decision-maker, or trained with sufficient data to achieve desirable accuracy, a machine can produce stable output without interference from cognitive overload, emotion, or other factors. In this case, machines typically support humans in a unilateral manner.

On the other hand, when the decision task is associated with higher uncertainty, the research opportunities become human-centered. High uncertainty makes many patterns in past data unaccountable due to inherent randomness in the risk event or human decision-makers' lack of understanding. Thus, the required complexity of algorithms increases to model and predict such data, and issues of overfitting and "black box" become vital (Topol, 2019; Amann et al., 2020). This challenge boils down to the demand for explainability, which means the machine should be able to explain its reasoning in a way that is understandable and trustable to humans (Cadario et al., 2021). In this regard, research on transparent AI, explainable AI, and trustable AI can help algorithms remove barriers that impede human decision-makers from valuing their output (Lyons and Havig, 2014; Gunning, 2016; Chen et al., 2018).

Furthermore, in decision tasks with high uncertainty, the research opportunities lie in human–machine collaboration centered for two reasons. First, humans are vulnerable to various cognitive biases, and their capabilities of information processing are limited, whereas machines can calibrate biases and handle mass data in a consistent and normatively correct way. When human and machine

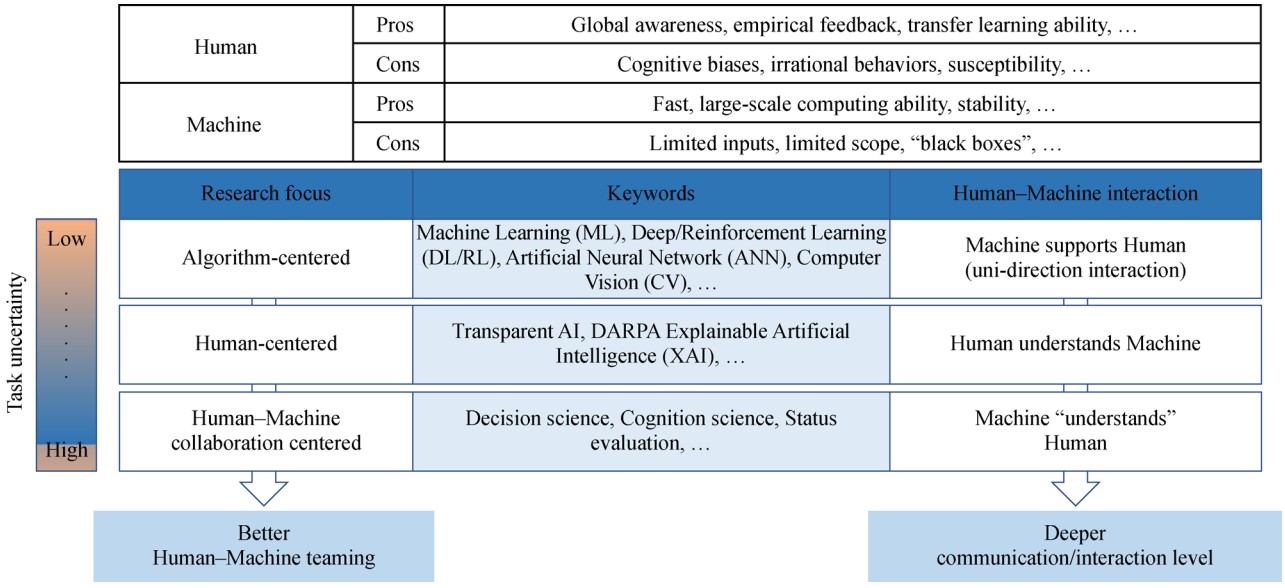

Fig. 2 Research opportunities on human–machine collaboration in risky decision-making.

judgments have disparity, machines need to be able to explain why the human judgments are normatively wrong, for example, by specifying which cognitive biases are taking effect, what information is lacking, which correlation/causality relationships are missing, and what regions of the solution space are underexplored. In this way, humans know that machines "understand" them, which helps cultivate trust through bilateral transparency and mutual understanding (Edmonds et al., 2019).

Second, machines are unable to handle highly uncertain and rare cases well. By contrast, humans can use intuition and experience to adapt to new situations and quickly learn and generalize reasoning across tasks. Representation of human cognitive and decision processes, for example, by abstract mathematical models or rule-based models, is an active topic in many fields such as cognitive science, psychology, behavioral science, and computer science (Renooij, 2001; Kemp and Tenenbaum, 2008; Broomell and Budescu, 2009; Tenenbaum et al., 2011). Understanding the sources of adaptability and generalizability in human judgments is key to equipping machines with capabilities to deal with new, complex, and rare cases.

Despite extensive literature on human–machine collaboration, research on human–machine collaboration (i.e., AI algorithm) in risky decision-making, especially about quantitatively assessing the likelihoods and consequences of each option and logically reaching a final decision, has drawn attention only recently. In the following section, we first summarize existing studies on human–machine collaboration, highlighting task allocation, human–machine relationship, and human–machine interaction, all of which are essential for decision tasks. We then discuss the challenges of human–machine collaboration in decision-makings under risk in the pursuit of high decision performance.

3 Literature on human–machine collaboration

We surveyed the literature on human–machine cooperation or collaboration from 1940 to 2021. Here, we do not limit the connotation of the machine. Instead, the machine could refer to an automated or autonomous system, an autonomous agent, a robot, an algorithm, or AI. Studies on human–machine collaboration span a variety of fields, including human–machine teaming (Calhoun et al., 2018; Daugherty and Wilson, 2018; Wynne and Lyons, 2018; Ferrari, 2019; Parker and Grote, 2019; Seeber et al., 2020; Laid et al., 2020; Saenz et al., 2020), human–machine relationship (de Visser et al., 2018; Lyons et al., 2018), transparency (Patel et al., 2019; Skraaning and Jamieson, 2019; Kraus et al., 2020), explainability (Gunning, 2016; Degani et al., 2017; DARPA, 2018; Amann et al., 2020; Cadario et al., 2021), task allocation (van Maanen and van Dongen, 2005; Roth et al., 2019; Dubois and Le Ny, 2020), acceptance (Gursoy et al., 2019; Shin, 2020), human trust in machine (Hoff and Bashir, 2015; de Visser et al., 2018; Gutzwiller and Reeder, 2021), (shared) mental models (Cannon-Bowers et al., 1993; Flemisch et al., 2012; Goodrich and Yi, 2013), situation awareness (Salmon et al., 2008; Ososky et al., 2012), measurement (Damacharla et al., 2018), and practice in decision-making (Jarrahi, 2018; Duan et al., 2019; Haesevoets et al., 2021).

Here we elaborate on current studies on human–machine collaboration from three perspectives, namely, the organizational perspective, the relationship perspective, and the interaction perspective (see Fig. 3). These perspectives correspond to different levels of deployment in human–machine collaboration, considering how humans and machines are organized, how they work together, and how they interact with each other. Specifically, the

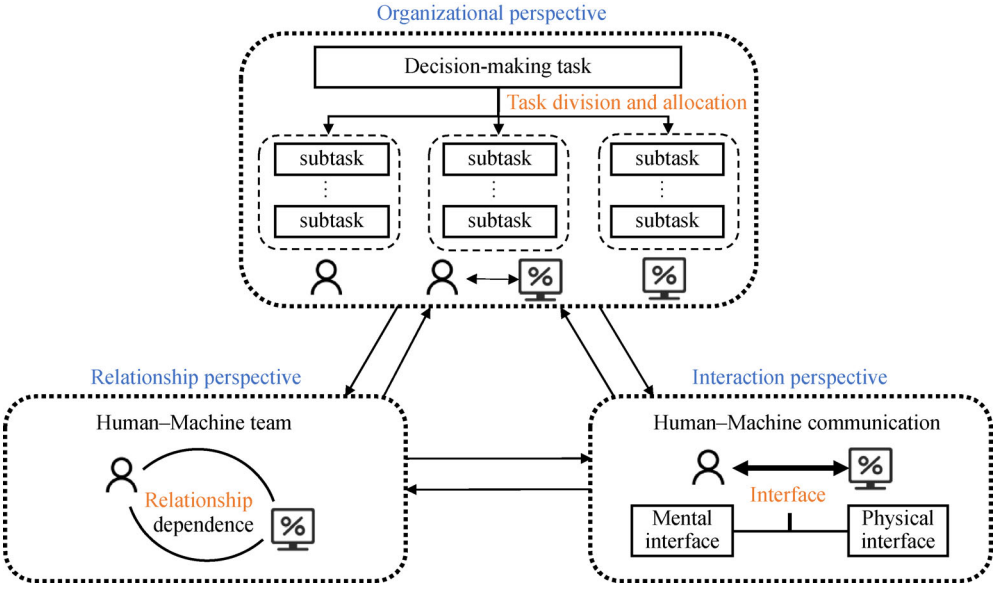

Fig. 3 Literature perspectives on human–machine collaboration.

organizational perspective concerns forming the human–machine team organizations and solving task allocation problems. The relationship perspective investigates acceptance, trust, and dependence of human and machine on each other. Finally, the interaction perspective is mainly about designs of communication to foster mutual understanding and bilateral interventions via physical and mental interfaces. We summarize studies taking each view in Tables 1–4, respectively.

3.1 Task allocation between human and machine in decision-making

Function allocation refers to distributing system functions and tasks across people and technology (Roth et al., 2019). The earlier representative work is theory of function allocation (Fitts, 1951). Functions are allocated according to the relative capabilities of humans and machines, that is, the "Fitts List" or "Men-are-better-at/Machines-are-better-at" classification scheme. This method divides and dedicates tasks to humans or machines according to their superior capabilities. Later, scholars argued that task allocation should stress cooperation and developed the theory of task allocation through the task and functional analysis (Chignell and Hancock, 1986). With the increase of automation levels, research attention gradually shifted to supervisory control. For example, Parasuraman et al. (2000) proposed the level of automation framework to specify the levels of automation for independent functions and the extent and process of human involvement.

As machine autonomy and intelligence improve, humans and machines are no longer analyzed separately but as a hybrid team (Chen and Barnes, 2014; Lyons et al., 2018; Wynne and Lyons, 2018; Lyn Paul et al., 2019). Findings in traditional task allocation studies are no longer sufficient to support effective human–machine teams with strong and intelligent machines. When assigning activities in human–machine teams, a "missing middle" emerges, that is, human–machine hybrid activities that are beyond human-only and machine-only activities (Daugherty and Wilson, 2018). A study by 65 scientists reported that collaboration design in human–machine teams is one of the

Table 1 Summary of studies on task allocation within human–machine teams

Topic	Main focus	Ref.
Task/Function allocation	Capability	Fitts (1951)
	Capability, task (definition, process)	Chignell and Hancock (1986); Parasuraman et al. (2000)
	Capability, task, human–machine team	Daugherty and Wilson (2018); Kunnathuvalappil Hariharan (2018); Karstens et al. (2018); Patel et al. (2019); Seeber et al. (2020); Tschandl et al. (2020)
	Task requirements, interdependence requirements	Roth et al. (2019)
	Openness of the problem, known risk level	Saenz et al. (2020)
Dynamic allocation	Dynamic characteristics (status, trust, workload, etc.)	Chignell and Hancock (1986); Lee and Moray (1992); Lee and See (2004); Wickens et al. (2013); Hancock et al. (2020); Dubois and Le Ny (2020)

Table 2 Summary of studies on human–machine relationship

Topic	Main focus	Ref.
Human–machine relationship	User and tool, pilot and co-pilot	Hoc (2000); Urlings and Jain (2002)
	Human-centered AI	Gunning (2016); de Visser et al. (2018); Li and Etchemendy (2018); Xu (2019)
	Machine as teammates	Phillips et al. (2011); Lyons et al. (2018); Wynne and Lyons (2018); Seeber et al. (2020)

Table 3 Summary of studies on the physical interface of human–machine interaction

Topic	Main focus	Ref.
Physical interface	Information display, the control of a machine/system	Fitts and Seeger (1953); Bradley (1954); Ortiz and Park (2011)
	Transparency, trust in automation	Seong and Bisantz (2008); Wickens et al. (2013); Hoff and Bashir (2015); Skraaning and Jamieson (2019)
	Transparent interface, mental model, effective human–machine team	Speier (2006); Kreye et al. (2012); Schaefer et al. (2017); Seeber et al. (2019); Ferrari (2019); Hancock et al. (2020); Gutzwiller and Reeder (2021)

Table 4 Summary of studies on the mental interface of human–machine interaction

Topic	Main focus	Ref.
Attitudes toward machine (attitudes)	Influencing factors, models and theories of acceptance	Fishbein and Ajzen (1975); Davis et al. (1989); Cramer et al. (2008); Kuo et al. (2009); Venkatesh et al. (2012); Gursoy et al. (2019); Du et al. (2019); Yalçın and DiPaola (2020)
	Influencing factors, models and frameworks of trust	Sheridan and Hennessy (1984); Lee and See (2004); McGuirl and Sarter (2006); Madhavan and Wiegmann (2007); Hancock et al. (2011); Hoff and Bashir (2015); Salem et al. (2015); Schaefer et al. (2016; 2017); Wang et al. (2016); Akash et al. (2017); Kraus et al. (2020)
Mental representation of machine (understanding)	Mental model, shared mental model	Cannon-Bowers et al. (1993); Johnson-Laird (1996); Gentner (2001); Vosgerau (2006); Kulesza et al. (2009); Ososky et al. (2012); Laid et al. (2020); Shin (2020)
	Situation awareness, shared situation awareness, situation awareness-based agent transparency models	Endsley (1988; 1995); Ososky et al. (2012); Selkowitz et al. (2016); Stowers et al. (2016); Chen et al. (2018); Bhardwaj et al. (2020)

three main design areas for human–machine collaboration (Seeber et al., 2020). Future human–machine teams should match tasks with the core capabilities of humans and machines. Roth et al. (2019) advocated integrated methods to support function allocation with high levels of autonomy considering work requirements and interdependence requirements for human–machine teams under routine and off-nominal (unexpected) conditions. Regarding decision-making tasks, in a recent study, Saenz et al. (2020) proposed that how humans and machines work together to make decisions depends on the openness of the problem (whether the problem is well defined and whether input variables are all known) and the risk level of the problem. For example, humans act as sentinels when faced with a relatively closed process with severe risks, whereas machines perform the task independently (sequential machine–human AI systems). By contrast, when the risk is low and the decision-making process is open, humans and machines cycle back and forth to reach the final decision and improve performance by mutual learning (cyclic machine–human AI systems).

Recent attempts have been investigating task allocation between humans and AI algorithms in several fields involving decisions under risk. For example, in financial planning, decisions are based on examining financial flows, anticipating the consequences of decisions, and weighing pros and cons (Kunnathuvalappil Hariharan, 2018). Kunnathuvalappil Hariharan (2018) discussed task allocation that leverages machine capabilities to collect information, discover links among multiple aspects, and picture an accurate outlook of prospective reactions to decisions, and human capabilities to identify critical factors and drill down to reach conclusions. In emergency decision-making, Karstens et al. (2018) developed a new human–machine mix paradigm to generate, inform, and utilize the information for predicting severe convective weathers, including four levels of automation and requiring human forecasters capable of transferring from one stage to another according to the evolution of hazard severity. In chest radiograph diagnosis, Patel et al. (2019) presented a practical case where the algorithm provides outputs for the presence of disease with confidence in probabilities while a human checks the outputs of lower confidence to achieve superior combined decisions. In skin cancer diagnosis, Tschandl et al. (2020) compared and discussed varied representations of AI-based supports in the clinical management decision and their effects across different levels of clinical expertise and situations. The researchers considered three strategies for dermatologist(s) and AI-based support to work together: Aggregating

AI-based multi-class probabilities and crowd wisdom, AI-based triage to screen patients, and using AI prediction as a second opinion for dermatologists' suspicious diagnoses.

Moreover, dynamic task allocation poses new challenges when the ways of human–machine interaction evolve over time. So far, studies have focused on dynamic task allocation considering human characteristics such as trust and cognitive load (Lee and Moray, 1992; Lee and See, 2004; Wickens et al., 2013). For example, with the advanced autonomous vehicle technology, vehicular control between the driver and the vehicle is allocated dynamically depending on driving circumstances and driver status (Hancock et al., 2020). Dubois and Le Ny (2020) developed a strategy based on the Markov decision process and quantitative models of trust and workload and showed the potential for improvement in human–machine collaborative decision-making for a general repeatable binary decision task.

3.2 Human–machine relationship

In earlier studies, humans and machines are regarded as users and tools rather than as a system or team (Phillips et al., 2011). For example, humans and machines collaborate as "pilot" and "co-pilot" respectively (Urlings and Jain, 2002), where the human is the ultimate decision-maker, and the machine acts as an automatic worker. Thus, the focus in these studies is on controllability and usability of machines (Hoc, 2000).

Given humans' fear of being replaced by superior machines, researchers have begun to advocate machine designs to follow the human-centered AI strategy (Li and Etchemendy, 2018; Xu, 2019). These studies emphasize that machines are meant to support and enhance humans rather than replacing them, in which machines are required to explain themselves to be understood. For example, the Explainable Artificial Intelligence project proposed by DARPA (DARPA, 2018) aims to make humans understand machines about, for example, why machines (algorithms) produce the current results, when they would succeed or fail, when they can be trusted, or why machines make mistakes (Gunning, 2016). Explainability and the explainable interface can promote human–machine collaboration and foster a teammate-like relationship (de Visser et al., 2018; Xu, 2019). In addition, Seeber et al. (2020) summarized that machine as teammates (MaT) requires not only machine artifact design but also collaboration design and institution design; hence, issues such as conversation, accountability, and task design would also affect the team collaboration. Moreover, the perception of task-independent relationship has been used as a dimension to describe teammate-likeness and would ultimately impact trust and team performance (Lyons et al., 2018; Wynne and Lyons, 2018).

3.3 Human–machine interaction

3.3.1 Physical interface

The physical interface is the physical channel of human–machine interaction, and its design focuses on reducing the gap in human–machine communication and promoting overall efficiency (Speier and Morris, 2003). The research on interface design considers different machine capabilities and roles in human–machine systems. In the era of mechanization, information display and interface layout are the main issues. The aim is to improve the operative control of the machine by helping people find the corresponding functions quickly, reducing the error rate of operation, and increasing efficiencies (Fitts and Seeger, 1953; Bradley, 1954; Ortiz and Park, 2011). In the automation era, many highly repetitive works are completed by machines, and humans become supervisors of the work process. Since then, the design of transparency to improve the usability and the control of complex systems has received extensive attention (Skraaning and Jamieson, 2019). In complex systems with high automation, such as nuclear power plants and aviation, a proper level of transparency is in demand to enable operators to understand the system strategy (Wickens et al., 2013) and the internal working conditions, leading to trust in automation (Hoff and Bashir, 2015; Schaefer et al., 2017) and appropriate usage (Seong and Bisantz, 2008).

The era of intelligence is characterized by massive information and a higher level of machine participation in human–machine collaboration. The physical interface pays more attention to human feelings and is expected to promote understanding and collaboration through transparent, effective human–machine interaction (Schaefer et al., 2017). Compared with non-iML, the behavior and judgment of iML are generally more credible and easier to interpret and thus can provide more effective assistance (Seeber et al., 2019; Gutzwiller and Reeder, 2021). Transparent interface with the representations of machine intent, perception of the environment, and system status can help establish the human mental model of machine and reduce the discrepancy between the human mental model and what the machine generates (Hancock and Chignell, 1989; Schaefer et al., 2017; Ferrari, 2019). Kreye et al. (2012) argued that human perception and judgment of uncertainty are subject to different displays of the same information and different contextual information. The representation of information is essential to the decision-making process and corresponding performance (Speier, 2006). For example, increasing the transparency of interface design can reduce decision-making conflicts between human and automated vehicles in autonomous driving (Hancock et al., 2020).

3.3.2 Mental interface

When machine capabilities of analysis, prediction, and decision grow, humans will have increasing difficulty understanding the "black box" of the machine. Lack of understanding impedes acceptance and trust. Therefore, in addition to the physical interface, how humans understand the machine (algorithm), including its behavior and mechanism, and what they think of or like/accept/trust the machine, entail a critical issue in human–machine collaboration (Chen et al., 2018; Cadario et al., 2021; Haesevoets et al., 2021). In this paper, we propose to use the term "mental interface" to describe the human's mental representation of the machine and attitudes toward the machine. The term "mental interface" was originally used in the field of computer science and engineering to refer to a brain–computer interface technology ("Cyberlink") that allows disabled persons with physical and/or mental disabilities to control a computer to play games and communicate at a basic level (Doherty et al., 2000; 2001).

To connect to the literature on psychology and human factors, the human's mental representation of the machine refers to the mental model (Gentner, 2001) and situation awareness (Endsley, 1988). Precisely, the mental model reflects the understanding of the surrounding environment (including the machine), whereas situation awareness depicts the human decision-maker's understanding of the task, especially in a dynamic context. Moreover, humans' attitudes toward the machine encompass acceptance (Davis et al., 1989) and trust (Lee and See, 2004).

Given close collaboration and significant differences between humans and machines, understanding and trust have come to the spotlight of research (Ososky et al., 2013), primarily affected by the human acceptance of machines. Low acceptance rates would discount the benefits of human–machine collaboration and team performance. Acceptance is related to a multitude of factors in human–machine collaboration, for example, machine capability (Gursoy et al., 2019; Yalçın and DiPaola, 2020), transparency (Cramer et al., 2008), explainability (Du et al., 2019), and human characteristics (Kuo et al., 2009). Acceptance is a research topic with a long history; hence, many acceptance models and theories can help predict people's acceptance level of machines and technology, with adaptations to AI (Fishbein and Ajzen, 1975; Davis et al., 1989; Venkatesh et al., 2012; Gursoy et al., 2019).

Moreover, to improve trust in human–machine teams, the two sides of decision-makers should understand each other and have a shared understanding of the task and environment (Cannon-Bowers et al., 1993; Ososky et al., 2012). This challenge requires establishing shared mental models and shared situation awareness through bilateral human–machine communication (Ososky et al., 2012; Chen et al., 2018). The (generalized) mental model is an explanatory model of how people understand the world, including the assumption, impression, and representation of themselves, teammates, organizations, and the world, which is limited by cognitive style and existing knowledge (Johnson-Laird, 1996; Gentner, 2001; Vosgerau, 2006). Like human teams, shared mental models in human–machine teams require mutual understanding and collective understanding of the environment. Machines should be aware of the human teammates' physical and mental capabilities and their intentions and motivations; whereas, humans should know the behaviors and working mechanisms of the machine teammates, including strengths, weaknesses, and especially their unusual behaviors and failure modes (Laid et al., 2020). The transparent and explainable decision-making process behind the machine, in turn, helps humans build their mental models, for example, correcting for cognitive biases and insufficient reasoning, supplementing missing information, and avoiding ignored risks (Kulesza et al., 2009; Shin, 2020). A shared situation awareness of the decision context and the environment (time and space), especially when they change dynamically, is another key to human–machine mutual understanding (Endsley, 1988; 1995). Shared situational awareness requires humans to understand machine status and intention. For example, the situation awareness-based agent transparency model (Selkowitz et al., 2016; Stowers et al., 2016; Bhardwaj et al., 2020) asks the machine to provide basic operation information for the human decision-maker to perceive, comprehend, and predict future events about the machine and the environment.

As a core element of the mental interface, trust plays the most pivotal role in human–machine collaboration. Trust is defined as the attitude toward the machine to help humans achieve the goal under the condition of uncertainty and vulnerability (Lee and See, 2004). In human–machine interaction, trust can influence people's acceptance and use of machines (Sheridan and Hennessy, 1984) and situation awareness and behaviors (Schaefer et al., 2017). Furthermore, mistrust or distrust weakens the effectiveness of human–machine teams (Hancock et al., 2011). As a multidimensional and dynamic construct, models or frameworks of trust have been widely studied and developed, including impact factors (Lee and See, 2004; Hancock et al., 2011; Hoff and Bashir, 2015; Salem et al., 2015; Schaefer et al., 2016), and trust dynamics and calibration (McGuirl and Sarter, 2006; Madhavan and Wiegmann, 2007; Wang et al., 2016; Akash et al., 2017; Kraus et al., 2020).

4 Challenges of human–machine collaboration in risky decision-making

On the basis of the discussion about critical components in human–machine collaboration for decision tasks, we

propose three challenges of human–machine collaboration in risky decision-making, pertaining to how to organize human–machine teams, enhance each other's capabilities, and facilitate mutual understanding and humans' trust in machines. In addition, we discuss difficulties and potential research directions.

Challenge 1: Developing a more dynamic and flexible human–machine team organization. Human–machine team organization plays a critical role in supporting the allocation of tasks and accountability between human and machine, affecting the performance of human and machine in the team (Flemisch et al., 2012). The challenge is specifically described below.

Designing the human–machine team organization mode to make decisions under risk. Humans and machines undertake different roles in the environment and tasks with different levels of variability, uncertainty, and complexity (Daugherty and Wilson, 2018). Authority seniority is usually determined beforehand and is consistent with the capabilities of human and machine decision-makers (Flemisch et al., 2012). However, humans and machines show variable capabilities in tasks and contexts with different levels of uncertainty. Thus, determining combinations of human and machine in human–machine teams and their seniority is challenging.

Applying dynamic task allocation strategy to respond to dynamic characteristics and to support the combined performance. For specific risky decision-making tasks, human–machine teams may encounter multiple environment uncertainty risk levels and exhibit dynamic behaviors (Bier et al., 1999). In this case, dynamic task allocation, based on the capability and characteristics of human and machine, can give better results than a static task allocation strategy (Dubois and Le Ny, 2020). However, the dynamic nature of risky decision-making tasks makes the situation unpredictable and poses higher requirements for the capability of both sides of human–machine teams, which induces difficulties for allocating tasks dynamically.

Determining appropriate accountability distribution in human–machine teams in risky decision-making. Risky decision-making is always accompanied by negative outcomes, and each stakeholder in human–machine teams has accountability for such outcomes (Shin and Park, 2019). An appropriate accountability distribution in a human–machine team can affect acceptance and facilitate a beneficial human–machine relationship (Flemisch et al., 2012). Human usually tends to blame the machine for the same mistake and negative outcomes (Dietvorst et al., 2015). This tendency would be more severe in risky decision-making with more uncertain negative outcomes. In addition, relevant laws and regulations are lacking, thereby posing challenges for supporting relevant research on accountability distribution.

To overcome **Challenge 1**, the following research questions must be considered.

(1) How should the human–machine team be organized and what are the criteria to decide which one (human, machine, or human–machine collaboration) holds the authority in risky decision-making?

(2) How should tasks between human and machine decision-makers, including cognition, judgment, and decision, be assigned? How can dynamic task allocation based on task requirements and the characteristics of human and machine decision-makers be achieved?

(3) What are the criteria to decide who should be accountable for the decision outcomes in human–machine teams in risky decision-making? How does different accountability distribution impact human–machine collaboration performance?

Challenge 2: Employing machines to help overcome humans' undesirable behaviors effectively (hence enhancing the human decision-maker) in risky decision-making. Existing studies pay more attention to how machines assist humans for better decision-making than to leveraging machines to discover and correct human cognitive and behavioral limitations in risky decision-making. We break down the challenge into three parts.

Determining the capability boundary of humans in risky decision-making. The capability boundary of humans is scoped by human cognitive and behavioral limitations in risky decision-making (Blumenthal-Barby and Krieger, 2015). Thus, understanding those limitations and their impacts is the basis of developing adaptive machines to overcome these deficiencies. Nevertheless, the context-dependent characteristic and insufficient research on the limitations in risky decision-making make the determination of the capability boundary difficult in specific risky decision-making situations.

Developing adaptive machine design to support in overcoming or intervening humans' multiple limitations. In risky decision-making, behaviors of human decision-makers, as well as multiple limitations in cognition and behavior, are affected by multiple dynamic and uncertain factors (Cokely and Kelley, 2009; Ordóñez et al., 2015). Therefore, an adaptive machine is requisite to help avoid negative outcomes in the case of rapidly changing behaviors. However, the variability of tasks and individual differences under the same task result in technical difficulties to develop corresponding machine functions to overcome human limitations.

Evaluating the collaborative decision-making process objectively and subjectively. Evaluating the collaborative decision-making process can help understand the collaborative process and move the machine design and human–machine collaborative design forward (Damacharla et al., 2018). Owing to the variability and uncertainty in risky decision-making, behaviors inside the human–machine team exhibit dynamic and complex characteristics. Besides, the collaboration process may shape human behaviors and cause long-term changes in relationships (Rahwan et al., 2019). These characteristics make a single indicator insufficient to evaluate the

complicated and evolving process. In addition, mere objective measurements cannot cover the dimensions of human–machine collaboration. Thus, subjective indicators, such as trust, comfort, and the capability to understand the human intention and to predict behaviors should be measured properly. Therefore, choosing indicators and measurement methods to evaluate the changes during the collaborative decision-making process is challenging.

To overcome **Challenge 2**, we raise the following research questions.

(4) What are human cognitive and behavioral limitations in risky decision-making? How can these limitations and their impacts be understood and modeled?

(5) How can machines provide normatively correct solutions for human cognitive and behavioral limitations? What impacts do different contexts or tasks have on human cognitive and behavioral limitations? In which way can machines be designed and developed to help overcome these limitations adaptively?

(6) What indicators can best describe and quantitatively evaluate the collaborative decision-making process?

Challenge 3: Developing communication and interface design to support mutual understanding and trust in human–machine teams. Communication and information sharing play a critical role in achieving an understanding of intentions and behaviors and creating an effective human–machine team (Chen et al., 2018; Edmonds et al., 2019). High uncertainty and task interdependence bring forward request to bi-directional transparency in real-time in human–machine teams (Gunning, 2016; Schaefer et al., 2017) to support effective communication and smooth task transition. More specific challenge details are described below.

Design for intention identification and alignment. The identification, understanding, and alignment of respective goal(s), value(s), and intention(s) in a human–machine team can improve the efficiency and performance of human–machine collaboration (Schaefer et al., 2017). However, human–machine teams do not have a common ground and linguistic interaction (Dafoe et al., 2021). Thus, problems would emerge when machines identify humans' explicit and implicit intentions in decision-making as well as signal their intentions.

Effective behavior identification and monitoring of behavioral limitations in risky decision-making. When the capability boundary is known, monitoring and identifying the human's irrational behaviors or behaviors due to cognitive limitations are critical for the intervention toward the human (Damacharla et al., 2018). Humans display various kinds of behaviors in risky decision-making. However, given the unclear relationship between used decision-making rules (cognition level) and observable behaviors, determining indicators from multiple observable behaviors creates difficulties for the real-time monitoring of the human's behaviors during the human–machine collaboration.

Appropriate intervention designs to overcome inconsistency in capabilities and behaviors in human–machine teams. When decisions of the human and machine decision-makers are inconsistent or capability/behavior limitations arise, appropriate intervention can effectively prevent possible negative outcomes (Daugherty and Wilson, 2018). However, given the nonlinguistic interaction, the machine has difficulty explaining its decisions and behaviors to humans. In addition, to implement effective intervention, a machine needs to overcome difficulties in understanding humans' decision-making rules and expressing them in an acceptable and understandable way.

Interaction design and evaluation considering human perception and understanding of machines. The physical interface has developed to be adaptive and algorithm dependent; more variables in the mental interface, such as trust and acceptance, should be considered to facilitate effective human–machine collaboration in risky decision-making (Dubois and Le Ny, 2020). The construction of mental interface depends heavily on embedding the model with relevant psychological variables, but models with these factors and their relationships have not been fully studied.

To overcome **Challenge 3**, we pose the following research questions for consideration.

(7) How do machines express their intentions, capabilities, and behaviors in risky decision-making? What behavioral indicators can represent human intentions? In which way can a human–machine team effectively align the goal, value, and intention?

(8) What behavioral indicators can represent the underlying cognition of human decision-making? How can machines identify and collect those indicators?

(9) How does a machine explain its decision-making rules? How does a machine understand humans' decision-making rules? How could the machine implement the intervention in an acceptable way?

(10) How can influencing factors in human–machine collaboration be modeled in risky decision-making? How can these models be embedded in algorithms behind the interaction interface?

To overcome those possible challenges (summarized in Fig. 4) and facilitate effective human–machine collaboration in risky decision-making, relevant disciplines and research show great necessity and potentials. Mutual enhancement is based on mutual understanding and the effective communication mechanism in human–machine teams and could be implemented through algorithms and interface designs. Therefore, technological advances in decision science, cognition science, and status evaluation, and the practice of integrating existing theoretical research into current human–machine teams can develop and extend human–machine collaboration in risky decision-making. Furthermore, the design of machine and interface comes from what the designer expects and the preset of the human–machine interaction mode and the

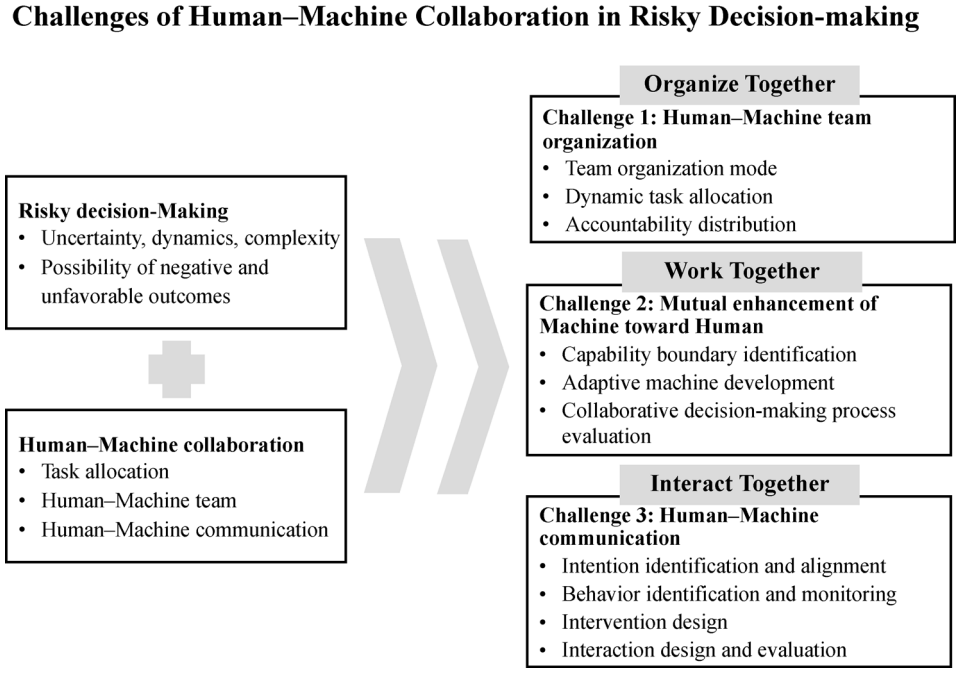

Fig. 4 Summary of challenges of human–machine collaboration in risky decision-making.

human–machine relationship. However, in reality, human and machine shape each other's behavior and gradually evolve into a stable and maybe different human–machine relationship. Thus, each part of research of human–machine collaboration in risky decision-making should be carried out from a dynamic and evolutionary perspective.

5 Conclusions

This paper focuses on challenges and opportunities in human–machine collaboration in risky decision-making, which is often characterized by uncertainty, complexity, and dynamics. We proposed three significant research challenges, covering how to organize efficient human–machine teams, foster healthy relationships between humans and machines, and build effective mental interfaces. We believe that this review can help researchers from multiple disciplines address exciting opportunities in this emerging field.

References

Akash K, Hu W L, Reid T, Jain N (2017). Dynamic modeling of trust in human–machine interactions. In: American Control Conference (ACC). Seattle, WA: IEEE, 1542–1548

Amann J, Blasimme A, Vayena E, Frey D, Madai V I (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20(1): 310

Apel H, Thieken A H, Merz B, Blöschl G (2004). Flood risk assessment and associated uncertainty. Natural Hazards and Earth System Sciences, 4(2): 295–308

Bedford T, Cooke R (2001). Probabilistic Risk Analysis: Foundations and Methods. Cambridge: Cambridge University Press

Bell D E (1982). Regret in decision making under uncertainty. Operations Research, 30(5): 961–981

Bhardwaj A, Ghasemi A H, Zheng Y, Febbo H, Jayakumar P, Ersal T, Stein J L, Gillespie R B (2020). Who's the boss? Arbitrating control authority between a human driver and automation system. Transportation Research Part F: Traffic Psychology and Behaviour, 68: 144–160

Bier V (2004). Implications of the research on expert overconfidence and dependence. Reliability Engineering & System Safety, 85(1–3): 321–329

Bier V M, Haimes Y Y, Lambert J H, Matalas N C, Zimmerman R (1999). A survey of approaches for assessing and managing the risk of extremes. Risk Analysis, 19(1): 83–94

Blumenthal-Barby J S, Krieger H (2015). Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. Medical Decision Making, 35(4): 539–557

Bradley J V (1954). Desirable control-display relationships for

moving-scale instruments. Technical Report 54–423. Dayton, OH: US Air Force, Wright Air Development Center (WADC)

Broomell S B, Budescu D V (2009). Why are experts correlated? Decomposing correlations between judges. Psychometrika, 74(3): 531–553

Cadario R, Longoni C, Morewedge C K (2021). Understanding, explaining, and utilizing medical artificial intelligence. Nature Human Behaviour, in press, doi:10.1038/s41562-021-01146-0

Calhoun G L, Ruff H A, Behymer K J, Frost E M (2018). Human-autonomy teaming interface design considerations for multi-unmanned vehicle control. Theoretical Issues in Ergonomics Science, 19(3): 321–352

Cannon-Bowers J A, Salas E, Converse S (1993). Shared mental models in expert team decision making. In: Castellan Jr N J, ed. Individual and Group Decision Making. New York: Taylor & Francis Psychology Press, 221–246

Charness G, Karni E, Levin D (2007). Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. Journal of Risk and Uncertainty, 35(2): 129–148

Chen G, Kim K A, Nofsinger J R, Rui O M (2007). Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. Journal of Behavioral Decision Making, 20(4): 425–451

Chen J Y C, Barnes M J (2014). Human–agent teaming for multirobot control: A review of human factors issues. IEEE Transactions on Human–Machine Systems, 44(1): 13–29

Chen J Y C, Lakhmani S G, Stowers K, Selkowitz A R, Wright J L, Barnes M (2018). Situation awareness-based agent transparency and human–autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science, 19(3): 259–282

Chignell M H, Hancock P A (1986). Knowledge-based load leveling and task allocation in human–machine systems. In: 21st Annual Conference on Manual Control. Moffett Field, CA: NASA Ames Research Center, 9

Cokely E T, Kelley C M (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. Judgment and Decision Making, 4(1): 20–33

Cramer H, Evers V, Ramlal S, van Someren M, Rutledge L, Stash N, Aroyo L, Wielinga B (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. User Modeling and User-Adapted Interaction, 18(5): 455–496

Croskerry P (2013). From mindless to mindful practice — Cognitive bias and clinical decision making. New England Journal of Medicine, 368(26): 2445–2448

Dafoe A, Bachrach Y, Hadfield G, Horvitz E, Larson K, Graepel T (2021). Cooperative AI: Machines must learn to find common ground. Nature, 593(7857): 33–36

Damacharla P, Javaid A Y, Gallimore J J, Devabhaktuni V K (2018). Common metrics to benchmark Human–Machine Teams (HMT): A review. IEEE Access, 6: 38637–38655

DARPA (2018). AI Next Campaign. Available at: darpa.mil/work-with-us/ai-next-campaign

Daugherty P R, Wilson H J (2018). Human + Machine: Reimagining Work in the Age of AI. Boston: Harvard Business Review Press

Davis F D, Bagozzi R P, Warshaw P R (1989). User acceptance of computer technology: A comparison of two theoretical models. Management Science, 35(8): 982–1003

Dawes R M, Faust D, Meehl P E (1989). Clinical versus actuarial judgment. Science, 243(4899): 1668–1674

de Visser E J, Pak R, Shaw T H (2018). From "automation" to "autonomy": The importance of trust repair in human–machine interaction. Ergonomics, 61(10): 1409–1427

Deck C, Jahedi S (2015). The effect of cognitive load on economic decision making: A survey and new experiments. European Economic Review, 78: 97–119

Degani A, Goldman C V, Deutsch O, Tsimhoni O (2017). On human–machine relations. Cognition Technology and Work, 19(2–3): 211–231

Dietvorst B J, Simmons J P, Massey C (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology, 144(1): 114–126

Doherty E, Cockton G, Bloor C, Benigno D (2001). Improving the performance of the cyberlink mental interface with the "Yes/No Program". In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM, 69–76

Doherty E, Stephenson G, Engel W (2000). Using a cyberlink mental interface for relaxation and controlling a robot. In: Proceedings of the SIGCAPH Computers and the Physically Handicapped. New York: ACM, 4–9

Dörner D, Wearing A J (1995). Complex problem solving: Toward a (computer simulated) theory. In: Frensch P A, Funke J, eds. Complex Problem Solving: The European Perspective. New York: Taylor & Francis Psychology Press, 65–99

Du N, Haspiel J, Zhang Q, Tilbury D, Pradhan A K, Yang X J, Robert Jr L P (2019). Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. Transportation Research Part C: Emerging Technologies, 104: 428–442

Duan Y, Edwards J S, Dwivedi Y K (2019). Artificial intelligence for decision making in the era of Big Data: Evolution, challenges and research agenda. International Journal of Information Management, 48: 63–71

Dubois C, Le Ny J (2020). Adaptive task allocation in human–machine teams with trust and workload cognitive models. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC). Toronto, ON, 3241–3246

Edmonds M, Gao F, Liu H, Xie X, Qi S, Rothrock B, Zhu Y X, Wu Y N, Lu H J, Zhu S C (2019). A tale of two explanations: Enhancing human trust by explaining robot behavior. Science Robotics, 4(37): eaay4663

Edwards W (1962). Subjective probabilities inferred from decisions. Psychological Review, 69(2): 109–135

El-Gamal M A, Grether D M (1995). Are people Bayesian? Uncovering behavioral strategies. Journal of the American Statistical Association, 90(432): 1137–1145

Endsley M R (1988). Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE National Aerospace and Electronics Conference. Dayton, OH, 789–795

Endsley M R (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1): 32–64

Ferrari V (2019). Man–machine teaming: Towards a new paradigm of

man–machine collaboration? In: Barbaroux P, ed. Disruptive Technology and Defence Innovation Ecosystems, vol. 5. Hoboken, NJ: John Wiley & Sons, 121–137

Fishbein M, Ajzen I (1975). Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Boston, MA: Addison-Wesley Publishing Company

Fitts P M (1951). Human Engineering for An Effective Air-Navigation and Traffic Control System. Washington, DC: National Research Council

Fitts P M, Seeger C M (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. Journal of Experimental Psychology, 46(3): 199–210

Flemisch F, Heesen M, Hesse T, Kelsch J, Schieben A, Beller J (2012). Towards a dynamic balance between humans and automation: Authority, ability, responsibility and control in shared and cooperative control situations. Cognition Technology and Work, 14(1): 3–18

Gentner D (2001). Mental models, psychology of. In: Smelser N J, Baltes P B, eds. International Encyclopedia of the Social & Behavioral Sciences. Amsterdam: Elsevier, 9683–9687

Goodrich M A, Yi D (2013). Toward task-based mental models of human–robot teaming: A Bayesian approach. In: International Conference on Virtual, Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments. Berlin, Heidelberg: Springer, 267–276

Gregory R, Slovic P, Flynn J (1996). Risk perceptions, stigma, and health policy. Health & Place, 2(4): 213–220

Grether D M (1992). Testing Bayes rule and the representativeness heuristic: Some experimental evidence. Journal of Economic Behavior & Organization, 17(1): 31–57

Griffiths T L, Tenenbaum J B (2006). Optimal predictions in everyday cognition. Psychological Science, 17(9): 767–773

Gunning D (2016). Explainable Artificial Intelligence (XAI) — What are we trying to do? Available at: cc.gatech.edu/~alanwags/DLAI2016/(Gunning) IJCAI-16 DLAI WS.pdf

Gursoy D, Chi O H, Lu L, Nunkoo R (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. International Journal of Information Management, 49: 157–169

Gutzwiller R S, Reeder J (2021). Dancing with algorithms: Interaction creates greater preference and trust in machine-learned behavior. Human Factors, 63(5): 854–867

Haesevoets T, de Cremer D, Dierckx K, van Hiel A (2021). Human–machine collaboration in managerial decision making. Computers in Human Behavior, 119: 106730

Hancock P A, Kajaks T, Caird J K, Chignell M H, Mizobuchi S, Burns P C, Feng J, Fernie G R, Lavallière M, Noy I Y, Redelmeier D A, Vrkljan B H (2020). Challenges to human drivers in increasingly automated vehicles. Human Factors, 62(2): 310–328

Hancock P A, Billings D R, Schaefer K E, Chen J Y C, de Visser E J, Parasuraman R (2011). A meta-analysis of factors affecting trust in human–robot interaction. Human Factors, 53(5): 517–527

Hancock P A, Chignell M H (1989). Intelligent Interfaces: Theory, Research and Design. North Holland: Elsevier Science Inc.

Hoc J M (2000). From human–machine interaction to human–machine cooperation. Ergonomics, 43(7): 833–843

Hoff K A, Bashir M (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3): 407–434

Holzinger A (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? Brain Informatics, 3(2): 119–131

Hunt R G, Krzystofiak F J, Meindl J R, Yousry A M (1989). Cognitive style and decision making. Organizational Behavior and Human Decision Processes, 44(3): 436–453

Jarrahi M H (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. Business Horizons, 61(4): 577–586

Johnson-Laird P (1996). Mental models, deductive reasoning, and the brain. In: Gazzaniga M S, ed. The Cognitive Neurosciences. Cambridge, MA: The MIT Press, 999–1008

Kahneman D, Frederick S (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In: Gilovich T, Griffin D, Kahneman D, eds. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge: Cambridge University Press, 49–81

Kahneman D, Tversky A (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2): 263–291

Karstens C D, Correia Jr J, LaDue D S, Wolfe J, Meyer T C, Harrison D R, Cintineo J L, Calhoun K M, Smith T M, Gerard A E, Rothfusz L P (2018). Development of a human–machine mix for forecasting severe convective events. Weather and Forecasting, 33(3): 715–737

Kemp C, Tenenbaum J B (2008). The discovery of structural form. Proceedings of the National Academy of Sciences of the United States of America, 105(31): 10687–10692

Kraus J, Scholz D, Stiegemeier D, Baumann M (2020). The more you know: Trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. Human Factors, 62(5): 718–736

Kreye M E, Goh Y M, Newnes L B, Goodwin P (2012). Approaches to displaying information to assist decisions under uncertainty. Omega, 40(6): 682–692

Kulesza T, Wong W K, Stumpf S, Perona S, White R, Burnett M M, Oberst I, Ko A J (2009). Fixing the program my computer learned: Barriers for end users, challenges for the machine. In: Proceedings of the 14th International Conference on Intelligent User Interfaces. Sanibel Island, FL: ACM, 187–196

Kunnathuvalappil Hariharan N (2018). Artificial Intelligence and human collaboration in financial planning. Journal of Emerging Technologies and Innovative Research, 5(7): 1348–1355

Kuo I H, Rabindran J M, Broadbent E, Lee Y I, Kerse N, Stafford R M Q, MacDonald B A (2009). Age and gender factors in user acceptance of healthcare robots. In: The 18th IEEE International Symposium on Robot and Human Interactive Communication. Toyama, 214–219

Laid J, Ranganath C, Gershman S (2020). Future directions in human machine teaming workshop. Arlington, VA: US Department of Defense

Lee J (2020). Is artificial intelligence better than human clinicians in predicting patient outcomes? Journal of Medical Internet Research, 22(8): e19918

Lee J, Moray N (1992). Trust, control strategies and allocation of function in human–machine systems. Ergonomics, 35(10): 1243–1270

Lee J D, See K A (2004). Trust in automation: Designing for appropriate

reliance. Human Factors, 46(1): 50–80

Li F F, Etchemendy J (2018). Introducing Stanford's human-centered AI initiative. Available at: hai.stanford.edu/news/introducing-stanfords-human-centered-ai-initiative

Luce R D, Fishburn P C (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. Journal of Risk and Uncertainty, 4(1): 29–59

Lyn Paul C, Blaha L M, Fallon C K, Gonzalez C, Gutzwiller R S (2019). Opportunities and challenges for human–machine teaming in cybersecurity operations. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63(1): 442–446

Lyons J B, Havig P R (2014). Transparency in a human–machine context: Approaches for fostering shared awareness/intent. In: International Conference on Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments. Cham: Springer, 181–190

Lyons J B, Mahoney S, Wynne K T, Roebke M A (2018). Viewing machines as teammates: A qualitative study. In: AAAI Spring Symposium Series. Palo Alto, CA, 166–170

Madhavan P, Wiegmann D A (2007). Similarities and differences between human–human and human–automation trust: An integrative review. Theoretical Issues in Ergonomics Science, 8(4): 277–301

March J G, Shapira Z (1987). Managerial perspectives on risk and risk taking. Management Science, 33(11): 1404–1418

McGuirl J M, Sarter N B (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. Human Factors, 48(4): 656–665

Mearman A (2011). Who do heterodox economists think they are? American Journal of Economics and Sociology, 70(2): 480–510

Miller A P (2018). Want less-biased decisions? Use algorithms. Harvard Business Review, 2018–7–26

Ordóñez L D, Benson III L, Pittarello A (2015). Time-pressure perception and decision making. In: Keren G, Wu G, eds. The Wiley Blackwell Handbook of Judgment and Decision Making, II. Hoboken, NJ: John Wiley & Sons, 517–542

Ortiz C A, Park M R (2011). Visual Controls: Applying Visual Management to the Factory. Boca Raton: Taylor & Francis Productivity Press

Ososky S, Schuster D, Jentsch F, Fiore S, Shumaker R, Lebiere C, Kurup U, Oh J, Stentz A (2012). The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. In: Proceedings of SPIE 8387, Unmanned Systems Technology XIV. Baltimore, MD, 838710

Ososky S, Schuster D, Phillips E, Jentsch F (2013). Building appropriate trust in human–robot teams. In: AAAI Spring Symposium: Trust and Autonomous Systems. Stanford, CA: Association for the Advancement of Artificial Intelligence, 60–65

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

Parker S, Grote G (2019). Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. Applied Psychology, in press, doi:10.1111/apps.12241

Patel B N, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, Rajpurkar P, Amrhein T, Gupta R, Halabi S, Langlotz C, Lo E, Mammarappallil J, Mariano A J, Riley G, Seekins J, Shen L, Zucker E, Lungren M P (2019). Human–machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digital Medicine, 2: 111

Payne J W, Bettman J R, Johnson E J (1993). The Adaptive Decision Maker. Cambridge: Cambridge University Press

Phillips E, Ososky S, Grove J, Jentsch F (2011). From tools to teammates: Toward the development of appropriate mental models for intelligent robots. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1): 1491–1495

Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon J F, Breazeal C, Crandall J W, Christakis N A, Couzin I D, Jackson M O, Jennings N R, Kamar E, Kloumann I M, Larochelle H, Lazer D, McElreath R, Mislove A, Parkes D C, Pentland A S, Roberts M E, Shariff A, Tenenbaum J B, Wellman M (2019). Machine behaviour. Nature, 568(7753): 477–486

Renooij S (2001). Probability elicitation for belief networks: Issues to consider. Knowledge Engineering Review, 16(3): 255–269

Roth E M, Sushereba C, Militello L G, Diiulio J, Ernst K (2019). Function allocation considerations in the era of human autonomy teaming. Journal of Cognitive Engineering and Decision Making, 13(4): 199–220

Saenz M J, Revilla E, Simón C (2020). Designing AI systems with human–machine teams. MIT Sloan Management Review, 61(3): 1–5

Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K (2015). Would you trust a (faulty) robot: Effects of error, task type and personality on human–robot cooperation and trust. In: 10th ACM/IEEE International Conference on Human–Robot Interaction. Portland, OR, 141–148

Salmon P M, Stanton N A, Walker G H, Baber C, Jenkins D P, McMaster R, Young M S (2008). What really is going on? Review of situation awareness models for individuals and teams. Theoretical Issues in Ergonomics Science, 9(4): 297–323

Schaefer K E, Chen J Y C, Szalma J L, Hancock P A (2016). A meta-analysis of factors influencing the development of trust in automation. Human Factors, 58(3): 377–400

Schaefer K E, Straub E R, Chen J Y C, Putney J, Evans III A W (2017). Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cognitive Systems Research, 46: 26–39

Seeber I, Bittner E, Briggs R O, de Vreede T, de Vreede G J, Elkins A, Maier R, Merz A B, Oeste-Reiß S, Randrup N, Schwabe G, Söllner M (2020). Machines as teammates: A research agenda on AI in team collaboration. Information & Management, 57(2): 103174

Seeber I, Waizenegger L, Seidel S, Morana S, Benbasat I, Lowry P B (2019). Reinventing collaboration with autonomous technology-based agents. In: Proceedings of the 27th European Conference on Information Systems (ECIS). Stockholm: Association for Information Systems, 4

Selkowitz A R, Lakhmani S G, Larios C N, Chen J Y C (2016). Agent transparency and the autonomous squad member. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1): 1319–1323

Seong Y, Bisantz A M (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. International Journal of Industrial Ergonomics, 38(7–8): 608–625

Sheridan T B, Hennessy R T (1984). Research and modeling of supervisory control behavior: Report of a workshop. Washington, DC: The National Academies Press, US National Research Council

Shin D (2020). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human–Computer Studies, 146: 102551

Shin D, Park Y J (2019). Role of fairness, accountability, and transparency in algorithmic affordance. Computers in Human Behavior, 98: 277–284

Silver D, Huang A, Maddison C J, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587): 484–489

Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T, Hassabis D (2017). Mastering the game of Go without human knowledge. Nature, 550(7676): 354–359

Simon D, Krawczyk D C, Holyoak K J (2004). Construction of preferences by constraint satisfaction. Psychological Science, 15(5): 331–336

Skraaning G, Jamieson G A (2019). Human performance benefits of the automation transparency design principle: Validation and variation. Human Factors, 63(3): 379–401

Speier C (2006). The influence of information presentation formats on complex task decision-making performance. International Journal of Human–Computer Studies, 64(11): 1115–1131

Speier C, Morris M G (2003). The influence of query interface design on decision-making performance. Management Information Systems Quarterly, 27(3): 397–423

Stowers K, Kasdaglis N, Newton O, Lakhmani S, Wohleber R, Chen J (2016). Intelligent agent transparency: The design and evaluation of an interface to facilitate human and intelligent agent collaboration. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1): 1706–1710

Tenenbaum J B, Kemp C, Griffiths T L, Goodman N D (2011). How to grow a mind: Statistics, structure, and abstraction. Science, 331 (6022): 1279–1285

Tetlock P E (2003). Thinking the unthinkable: Sacred values and taboo cognitions. Trends in Cognitive Sciences, 7(7): 320–324

Tong J, Feiler D (2017). A behavioral model of forecasting: Naive statistics on mental samples. Management Science, 63(11): 3609–3627

Topol E J (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1): 44–56

Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Janda M, Lallas A, Longo C, Malvehy J, Paoli J, Puig S, Rosendahl C, Soyer H P, Zalaudek I, Kittler H (2020). Human–computer collaboration for skin cancer recognition. Nature Medicine, 26(8): 1229–1234

Tversky A, Kahneman D (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157): 1124–1131

Urlings P, Jain L C (2002). Teaming human and machine: A conceptual framework. In: Abraham A, Köppen M, eds. Hybrid Information Systems. Heidelberg: Springer, 711–721

van Maanen P P, van Dongen K (2005). Towards task allocation decision support by means of cognitive modeling of trust. In: Proceedings of 17th Belgian-Netherlands Artificial Intelligence Conference. Brussels, 399–400

Venkatesh V, Thong J Y L, Xu X (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. Management Information Systems Quarterly, 36(1): 157–178

von Neumann J, Morgenstern O (1944). Theory of Games and Economic Behavior. Princeton: Princeton University Press

Vosgerau G (2006). The perceptual nature of mental models. Advances in Psychology, 138: 255–275

Wakker P (1989). Continuous subjective expected utility with non-additive probabilities. Journal of Mathematical Economics, 18(1): 1–27

Wang N, Pynadath D V, Hill S G (2016). Trust calibration within a human–robot team: Comparing automatically generated explanations. In: The 11th ACM/IEEE International Conference on Human–Robot Interaction. Christchurch, 109–116

Warden T, Carayon P, Roth E M, Chen J, Clancey W J, Hoffman R, Steinberg M L (2019). The national academies board on human system integration (BOHSI) panel: Explainable AI, system transparency, and human machine teaming. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63(1): 631–635

Whelehan D F, Conlon K C, Ridgway P F (2020). Medicine and heuristics: Cognitive biases and medical decision-making. Irish Journal of Medical Science, 189(4): 1477–1484

Wickens C D, Hollands J G, Banbury S, Parasuraman R (2013). Engineering Psychology and Human Performance, 4th ed. New York: Taylor & Francis Psychology Press

Wickham P A (2003). The representativeness heuristic in judgements involving entrepreneurial success and failure. Management Decision, 41(2): 156–167

Wynne K T, Lyons J B (2018). An integrative model of autonomous agent teammate-likeness. Theoretical Issues in Ergonomics Science, 19(3): 353–374

Xu W (2019). Towards human-centered AI: A perspective from human–computer interaction. Interaction, 26(4): 42–46

Yalçın Ö N, DiPaola S (2020). Modeling empathy: Building a link between affective and cognitive processes. Artificial Intelligence Review, 53(4): 2983–3006

Zinn J O (2008). Heading into the unknown: Everyday strategies for managing risk and uncertainty. Health Risk & Society, 10(5): 439–450