Bidirectional Communication for Effective Human-Agent Teaming

Amar R. Marathe(✉), Kristin E. Schaefer, Arthur W. Evans, and Jason S. Metcalfe

Human Research and Engineering Directorate,
United States Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA
{amar.r.marathe.civ,kristin.e.schaefer-lay.civ,
arthur.w.evans20.civ,jason.s.metcalfe2.civ}@mail.mil

Abstract. The recent proliferation of artificial intelligence research is reaching a point where machines are able to learn and adapt to dynamically make decisions independently or in collaboration with human team members. With such technological advancements on the horizon, there will come a mandate to develop techniques to deploy effective human-agent teams. One key challenge to the development of effective teaming has been enabling a shared, dynamic understanding of mission space, and a basic knowledge about the states and intents other teammates. Bidirectional communication is an approach that fosters communication between human and intelligent agents to improve mutual understanding and enable effective task coordination. This session focuses on current research and scientific gaps in three areas necessary to advance the field of bidirectional communication between human and intelligent agent team members. First, intelligent agents must be capable of understanding the state and intent of the human team member. Second, human team members must be capable of understanding the capabilities and intent of the intelligent agent. Finally, in order for the entire system to work, systems must effectively integrate information from and coordinate behaviors across all team members. The combination of these three areas will enable future human-agent teams to develop a shared understanding of the environment as well as a mutual understanding of each other, thereby enabling truly collaborative human-agent teams.

Keywords: Automation · Autonomy · Robot · Mixed-initiative teams
Human automation interaction · Bidirectional communication

1 Introduction

Integration of advancing intelligent technologies on the battlefield will change the very nature of the tasks Warfighters need to perform. This, in turn, will require the evolution of different skill sets and capabilities, which will thus impact the precise needs for those Warfighters. A research and development approach is needed that conceives of not only the potential capabilities of these future intelligent technologies, but the potential for completely novel interactions among heterogeneous teams of Warfighters and intelligent agents, and reconceives approaches and requirements for training.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 338–350, 2018.
https://doi.org/10.1007/978-3-319-91581-4_25

Within these teams, the human is integral to decision-making, including adapting requirements to dynamic events, and completion of the overall mission. As such, there is a need to have the human-agent team perform as well as human-only teams but with the potential to provide additional support for advanced mission directives. The benefits of integrating intelligent agents into a human team include the potential for greater team resilience with robust, adaptive performance; faster, dynamic human-agent teaming (HAT) reconfiguration to match capabilities to mission requirements; faster, more informed team decision making; and reduced numbers and risk to human team members.

Developing advanced intelligent technologies that are capable of functioning as a natural teammate is a critical challenge for the research community. Research into successful human teams has shown that performance outcomes are not simply a sum or an average of the performance of the individuals. Instead, emergent properties are the result of the interaction of the components of the team, which cannot be reduced to or described wholly in terms of the individual elements of the system considered in isolation. Team performance often breaks down because of problems with emergent team states and process such as insufficient communications, misunderstanding of team goals, undefined team responsibilities or lack of shared mental models, and conflict [1, 2].

Moving from human-only teams, to mixed-agent teams composed of humans, intelligent software agents, embodied agents (e.g., robots), and networked sensors adds complexity that may not be completely comprehended today; this especially in uncertain informational environments with limited or unreliable communications. To develop effective mixed agent teams, humans and agents must be allowed to work in disparate dimensions (time, space, world views, representations, mental models, etc.), but also capable of seamlessly synchronizing for collective action. For example, intelligent agents will process information, reason, and make decisions at scales beyond that of humans in both time and scope, and yet we will want to include humans in the decision-making loop for many if not most battlefield decisions on account of their superior abilities to adapt and develop abstract understandings in the face of novelty and uncertainty. This brings to light the open research question of how to capitalize on the individual advantages of both humans and agents, and simultaneously enhance the performance of the collective group.

While this is largely an open question, part of the solution to these issues is to developing methodologies to enable bidirectional communication between the human and intelligent agents. Successful human teams often can communicate effectively, have team members who possess knowledge of each other's skills and capabilities that allow them to anticipate each other's actions, and can interpret and assess the environmental constraints on the task at hand. Building a robust capability for resilient bidirectional communication is believed to be an important approach to developing these same properties in human-agent teams by facilitating increased mutual understanding between human and intelligent agents. In this session, we discuss three broad challenge areas that must be addressed to enable the development highly functioning human-agent teams.

Agents Understanding Humans. Intelligent agents must be provided with information necessary to develop an understanding of their human team members. This requires leveraging wearable (and non-wearable) technologies to develop human assessment tools capable of generating high-resolution, real-time, predictions of an individual's internal and external behavioral and performance dynamics across a variety of environments. The specific types of predictions necessary will depend on how that information is to be used, the current task, and the current environment. Two uses of the information are considered in the sections that follow.

Humans Understanding Agents. Humans must be capable of understanding the capabilities and intent of their intelligent agent teammates. This requires developing adaptive interfaces for the intelligent agents that provide the human with the appropriate level of detail regarding the intelligent agent decisions or behavior. These adaptive interfaces will leverage human assessment tools described in the previous section to customize the displays for the current state of a particular human. These displays must convey information about the intelligent agent behavior or decisions with consideration for issues such as transparency, trust, and team situation awareness.

Joint Human-Agent Teamwork. Integration of intelligent agents and Warfighters requires the capability to deliver appropriate information to both the human and intelligent agent at the appropriate time. This necessitates the need for developing integration principles and approaches that dynamically accentuate the strengths of individual humans and agents while mitigating relative weaknesses for improved performance. This type of integration requires insight into the current and future states of each individual agent, which, for humans, should come from the human assessment tools described in the Sect. 2.

2 Agents Understanding Humans

Human-agent team performance is not limited by computing power, but by the ability for computer or embodied intelligent agents to understand humans. This is evidenced by overwhelming majority of current systems that assume fixed, stereotyped human input that is geared toward an "average" human. These systems assume that the quality of human input, which is often presumed to be of low noise and high accuracy, will be static over time and across individuals. We consider this to be a fundamentally flawed assumption because of inter and intra-individual differences in humans and agents. To account for human variability during HAT, continuous, real-time human assessment technologies, which will combine human and environmental sensing technologies with advanced analytics, will provide the foundational elements for future systems to generate high-resolution, moment-to-moment, predictions of individual's internal and external behavioral and performance dynamics across training and operational environments. Fundamentally, this capability will enable future systems to move from an approach of *mitigating* the effects of human variability to one that *embraces and predictively capitalizes* on that variability.

The US Department of Defense, Defense Science Board [3] report on autonomy suggests that the future value of unmanned systems lies not in the direct replacement of any one particular human operator, but rather in their contribution to overall human-system collaboration (e.g., the capability to extend and complement human capability without degradation due to factors such as fatigue, stress, lack of attention, situation awareness, amongst others). Further, calibrating inter-agent trust is considered essential to this collaboration. Some identified critical gaps to developing trust in HAT include the impact of human states (stress, fatigue, and attentional control), cognitive factors (understanding of technology, ability to use or interact with technology and expectance), and emotive factors (confidence, attitudes, satisfaction and comfort) on teaming with respect to task-specific environmental factors including risk, uncertainty, task type, context, and the physical environment [4]. While a number of current research efforts are underway to better understand and quantify these critical gaps, continued research will advance our understanding of the human performance during real-time, real-world operations, as well as advance adaptive autonomy technologies leading to more advanced, collaborative teaming.

Near and mid-term research will focus on leveraging sensing technologies to enable high fidelity, omnipresent prediction of behaviors and intentions that can account for continuous changes in an individual's physical, cognitive, and social states (examples from these three types of states include stress, workload, fatigue, task difficulty, trust, and situation awareness). The goal is to enable the exploitation of the array of sensors and information streams to predict human performance dynamics with sufficient resolution and robustness to adapt systems in manners to directly enhance performance.

The current state of the field has demonstrated unparalleled advancements in sensor and analysis technologies that provide new insights into different facets of human psychology, physiology, behavior, and performance. For example, advances in neuroscientific tools have revealed novel discoveries on how differences in brain function influence precise human behaviors [5, 6]. Advances in social and environmental sensing tools and have provided unprecedented insights into patterns of gross human social behaviors [7], while advances in biochemical or fluid sensing (i.e., blood, sweat, and tears) are providing unique insights into the continuous dynamics of internal human states and traits. More generally, advances in wearable devices have enabled the tracking of a wide range of factors including activity, sleep patterns, and various physiological parameters (see [8] for an overview).

Even with this broad explosion in sensing technologies, there is a lack of understanding regarding the factors that influence variability in human performance, as well as an inability to develop predictive algorithms that account for this variability. This has prevented a similar explosion in human assessment technologies capable of providing robust predictions regarding health and performance. As we develop a better understanding of human variability, a combination of sensing, analytic, and enabling technologies must be leveraged to assess and predict human states, behaviors, intentions, and performance. This capability will provide the foundation for a broad range of individualized, adaptive technologies across a variety of domains. Understanding the human team member in such detail can then be used to influence autonomy design. A sampling of critical technologies include:

- Critical Sensing Technologies, including wearable and non-wearable sensors, provide novel insights into the human and their local environment that could be communicated to an intelligent agent. These types of sensors provide data regarding intended behaviors, unintended behaviors, physiology, brain activity, subjective experiences, and social interactions; as well as task constraints and a wide range of environmental and societal factors. These sensing technologies arise from fields including quantified self; brain-computer interface; human augmentation; bioacoustic sensing; electrovibration; speech recognition and translation; gamification; wearable user interfaces; mobile health monitoring; gesture control; activity streams; and pervasive computing. The sensors fall in the following general categories of on-body sensing (e.g., watches, shirts, cell phones, glasses), off-body sensing (e.g., cameras, computers, fixed-placement), and their combinations for use in network-based sensing (e.g., network interactions, team and organizational performance, societal events).

- Critical Analytic Technologies merge and interpret data from a wide variety of human, task, and environmental sources to communicate higher-level goals and understanding of the mission space and human team members. These arise from fields including: brain-computer interface, affective computing, prescriptive analytics, natural-language response, big data, complex-event processing, content analytics, location intelligence, social network dynamics, and predictive analytics.
- Critical Enabling Technologies provide an integration framework for consolidating these data into a tractable repository for analysis by augmenting current network and computational technologies. Examples of these technologies will rely upon wireless local area networking, wireless bridge technologies, frequency domain equalization optimization, cloud computing, and smart antennae. These types of technologies provide fundamental solutions for collecting, storing and analyzing sensed data in real-time, thereby enabling predictions of human state and performance to be conveyed to intelligent agents.

3 Humans Understanding Agents

Intelligent technologies, such as embodied agents, are not yet operating as teammates in the field. They are by and large either teleoperated or supervised tools that possess insufficient shared understanding to independently adapt to the benefit of the team [9]. However research efforts are underway to advance intelligence architectures for independent and collaborative decision-making capabilities that can account for uncertain and dynamic environments [10, 11]. As agents become more sophisticated and independent, it is critical for their human counterparts to understand their behaviors, the reasoning process behind those behaviors, and the expected outcomes to properly calibrate their trust in the systems and make appropriate decisions [12, 13]. This is essential to the team effort because people will question the accuracy and effectiveness of agents' actions if they have difficulty understanding the state or status of the agent [14–16] and the reasoning behind specific actions or behaviors [17–19]. As such, if user expectations do not match agent actions (even if the actions of the agent are optimal and appropriate decisions), then trust can degrade to the degree in which the user will misuse or disuse the agent [20, 21].

These limitations with how humans understand agents can be substantial impediments to overall system, task, and team performance. To overcome these limitations, future systems will leverage detailed assessments of the human, as described in Sect. 2, to dynamically adapt and tailor the interface to fit the informational needs dictated by the precise teaming function or functions to be shared between the autonomous agent and human. These adaptive interfaces will ensure that the human is given the information needed at the appropriate time and in the appropriate scale and frame of reference to maximize overall performance.

Recent research efforts examine the integration of spatial and temporal context into artificial intelligence (AI) development. Semantic mapping is used to label objects in the world in order to assign high-level information to decisions, such as those needed for communicating decision related to path generation and mapping [22, 23]. Computer vision [24, 25] and natural language [26] have been used to develop probabilistic models that can provide an abstraction of the environment and better support intelligent agent-to-human communication needs. Research into temporal context and AI has looked at time perception in decision-making [27]. This work is important because it helps in the development of shared situation awareness by enabling the capability of interpreting the state or recognizing the current situation by observing the partial or entire state history at any time.

This link between AI and the human's understanding of the agent can directly influence bidirectional communication needed for appropriate trust development through what Lee and See [12] termed the 3 P's: purpose, process, and performance. Purpose deals with the degree to which the agent-driven automation is being used according to the designer's intent. Process deals with the question of whether the algorithm of the automated system is appropriate for a given task. Performance deals with system reliability, predictability, and capability. Lee [28] proposed that to increase system transparency to the human team member, the system's 3Ps, as well as the history of the 3Ps, should be visible to the operator. However, the presentation should be in a simplified form (e.g., integrated graphical displays) so the operator is not overwhelmed by the additional information he/she needs to process [29–31]. From this basis, Chen et al. [18] developed the Situation awareness-based Agent Transparency (SAT) model, which identifies and organizes the information that an agent needs to share with the human teammate to support their situation awareness, trust, and appropriate reliance upon the agent. This effort demonstrates the potential for developing interfaces that dynamically adapt to the state and needs of the team members (possibly quantified by wearable technologies), the task at hand, and the situational context to provide necessary and sufficient explanations to enhance team performance while simultaneously building and maintaining team cohesion.

As an extension to the transparency issue, trust in automation (TiA) has long been considered central in influencing the way a human user interacts with an automation; if TiA is too high there will be overuse, if TiA is too low there will be disuse [12]. TiA is an important construct that undoubtedly affects human user interaction behaviors. However the relationship between TiA and human behavior is complex and not currently fully understood. Further, relevant to immediate real-world applications for improving team performance, TiA measurement has most commonly leveraged subjective metrics [32] occurring after an interaction, with only some more recent research efforts looking at behavioral indicators and physiological markers of trust to map to potential real-time implications [13, 21, 33–35]. However, current research has suggested that transparent communication of agent intent can support collaboration and in turn calibrate trust and reliance on the system [36]. The specific criteria for this communication are still being developed.

Further, from an engineering standpoint, it is not yet possible to define an objective function based on TiA that would adequately define how control authority should shift dynamically between operator and automation. By contrast, specific behaviors are readily observed and measured in real time and do not have the confounding effect of inferring psychological causality. Recent studies have shown that an operator's decision to toggle control between human and autonomy drivers may be predicted on the basis of behavior, physiological, and environmental factors [37]. These predictions can integrated with dynamic estimates of performance to calibrate trust-based decisions [35].

As the future of human-agent interaction moves towards interdependent teaming initiatives, developing efficient complex decision-making processes is an essential part of the design and development of autonomous agents. The process by which agents make decisions is still an open research question when operating in uncertain environments with unknown data. Current research has demonstrated the importance and relevance to why understanding agent decision-making processes is relevant to performance. It has been shown that humans and intelligent-agents faced with the same circumstances will not make the same decisions, even under the same set of apparent constraints, nor will they necessarily have the same consequences resulting from those decisions. Perelman and colleagues [38] found that when comparing human and algorithmic path planning, there is more than one 'human way' of solving a planning problem which may or may not match an algorithm. The potential mismatch of solutions, without explanation, could result in significant degradation of team process. Human teammates may reduce their interaction with or outright ignore intelligent agents regardless of how correct the solutions are which they provide.

4 Joint Human-Agent Teamwork

In order to develop effective mixed-agent teams, technologies for inferring motivations, predicting behavior and reasoning about the environment must be incorporated into a closed-loop system that can initiate individualized interventions to improve team performance. In general, humans are able to adapt to the complexities and dynamics of real-world operational environments to a degree unmatched by current forms of autonomy. However, humans simply cannot process the full amount of information available or understand the reasoning of complex, intelligent agents. Thus, we must develop novel integration principles and approaches that leverage advanced sensing, data processing, and dynamic inferential tools in a way that can enable us to accentuate the strengths of individual humans and agents while mitigating relative weaknesses for improved decision making.

It has long been understood that, though automation can execute predictable, well-defined procedures with superior speed and reliability, humans are far superior at tasks that require inductive reasoning and adaptation to novel or changing information [39–

41]. As a result, system integrators have developed a wide range of approaches to supplement intelligent agent autonomy with human inputs to increase resilient and robust performance within complex, dynamic, and uncertain environments. Yet most systems-level, human-centric design approaches have treated the human as most appropriately positioned at the peak of the command hierarchy (c.f. [42–45]) rather than as a fully collaborative teammate [46]. That is, while these approaches have not always required the user to give continuous control or decision inputs and corrective feedback, when the human input has been available, it has usually been integrated as the *de facto* correct solution, which is not always true. Treating the human as the ultimate and final authority is a premise based on either an explicit mandate [42] or on an assumption that human influence would guarantee optimal behavior in situations where an automated agent is uncertain or otherwise compromised.

An ethical debate has emerged as to whether or not a person should retain overarching decision-making authority and perhaps more importantly, accountability [47]. Are there times that an intelligent-agent should make the decision rather than the user? We do not see the answer to this question as black-and-white. Variables such as context and limitations associated with the intelligent-agent's ability to physically detect and make sense of the environment, as well as to infer the intent of the human teammate, will critically impact whether a response is appropriate or inappropriate. Human intervention may be appropriate in one condition and inappropriate in another. We acknowledge the complex and sometimes emotional aspects of this debate. This transfer process is difficult to manage and requires balancing control allocation within the team structure while maintaining team commitments and supporting large-scale interactions with multiple agents. Reengagement for a person can sometimes be difficult and may impact safety and mission effectiveness due to a user's situation awareness, workload, and abilities. We join those who have argued that adherence to this premise has limited how well human inputs have been integrated with autonomous systems [46, 48, 49].

There has been considerable research into mitigating the potential impact of human variability and performance failures on HAI systems. Unfortunately, the majority of these approaches have only succeeded in limited and controlled contexts, and have not been widely adopted for real-world use. We argue that this is due, at least in part, to adherence to the axiomatic premise that the human should be the ultimate and final authority; with failure to fully account for the dynamic strengths and vulnerabilities of the human team member being a critical design outcome of this belief.

Recent efforts have proposed the Privileged Sensing Framework (PSF), an evolved approach that treats the human as a special class of sensor rather than as the absolute command arbiter [50]. This approach is based on the concept of appropriately 'privileging' information during the process of integrating information from human and autonomy team members by bestowing advantages, special rights, or immunities based on the characteristics of each individual agent (on the basis of data from wearable and non-wearable sensors), the task context, and/or the performance goals. Indeed, treating the human as a privileged sensor deviates from the established central axiom of human-centered automation [51]. However, we view this departure as an important evolutionary step beyond substitution-based function allocation methods [52] and in alignment with notions of human-automation interactions that capture a more authentic essence of natural teaming behavior [46, 53, 54].

5 Session Details

Focusing efforts on developing bidirectional communication as a critical capability may be an essential approach to human-agent teaming that seeks to develop common ground and shared understanding between human and intelligent agent team members. In this session, we discuss current research and gaps within three focus areas. First, intelligent agents must be capable of understanding the state and intent of the human team member. This area of research is largely focused on sensing the human and providing the intelligent agents with information related to estimated state and expected performance of that human team member. Second, human team members must be capable of understanding the intelligent agent. This area of research is focused on conveying information about the intelligent agent back to the human team member in a manner most appropriate to that human team member at a given point in time. This means, that many of these approaches will involve adapting the information conveyed to the human based on the information the agent has about that team members current state. Finally, in order for the entire system to work, we must develop closed loop solutions to effectively integrate information from and coordinate behaviors across human and intelligent agent team members. This area of research focuses on developing integration principles and approaches that accentuate the strengths of individual humans and agents while mitigating relative weaknesses for improved team performance. Within this session, we have three talks that cover a subset of these areas.

5.1 A Maximum Likelihood Method for Estimating Performance

Jonroy Canady and colleagues provide an example of estimating human decisions that may be used to improve human agent teaming within the context of a joint human-agent target identification paradigm [55]. In this work, they present a method for creating unbiased and accurate estimates of human target-detection performance that could be used to better integrate the human decisions with those of the autonomous agents. This study provides an example of how information sensed from human responses can be conveyed to an intelligent agent. Within the bidirectional communication framework, an intelligent agent teammate would then leverage that information to adapt their behavior in a manner that improves overall system performance.

5.2 Quantifying Human Decision-Making

Kristin Schaefer and colleagues address the importance of human spatial decision-making as it applies to the development of appropriate human-agent team communication [56]. In this work, they aim to quantifying human spatial decision-making because if human behavior does not match the robots' models or expectations, there can be a degradation in trust that can impede team performance. Degradation in trust may only be mitigated through explicit communication which are needed to develop common ground and a shared understanding. To reach this end, this study identifies divergence in planning and action times and behaviors as well as detects differences in local versus global decision-making processes needed to predict complex decisions within the confines of increasing environmental complexity and amount of prior knowledge about the task. It supports agents understanding humans in that findings from quantifying human spatial decision-making can advance the technical capabilities of a robot to more accurately perceive and interpret human team member behavior. The next step in the research will support humans understanding agents as these findings are compared to robot solutions. Where disparities exist between the resultant robot and human behaviors, bidirectional communication can be used as a means to achieve an optimal solution collaboratively.

5.3 The Role of Psychophysiological Measures Within Mixed-Initiative Teams

Kim Drnec and colleagues extend their previous work on characterizing behavioral, physiological and task-based factors that influence trust-in-autonomy by examining the efficacy of a real-time system that uses a combination of behavioral and psychophysiological data from a human driver to foster more appropriate use of autonomous driver assist technologies [57]. This study provides an example of a system using bidirectional communication between human and intelligent agent to improve overall performance. The system uses information from wearable sensors about behavior, and physiology and combines that with environmental and task based factors to predict human intent to toggle control (*Agent Understanding Warfighter*). The system then uses that information about the human driver to provide customized feedback to assist the driver in making an appropriate decision regarding whether or not to toggle driving control (*Warfighter Understanding Agent*). The overall system also shows a simple case of using bidirectional communication to share information between a human and intelligent agent to maximize performance.

Acknowledgements. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Donaldson, M.S., Corrigan, J.M., Kohn, L.T., et al.: To Err is Human: Building A Safer Health System. National Academies Press, Washington D.C. (2000)
2. Salas, E., Stagl, K.C., Burke, C.S., Goodwin, G.F.: Fostering team effectiveness in organizations: toward an integrative theoretical framework. In: Nebraska Symposium on Motivation, p. 185 (2007)
3. Murphy, R., Shields, J.: The role of autonomy in DoD systems. Def. Sci. Board (2012)

4. Schaefer, K.E., Chen, J.Y., Szalma, J.L., Hancock, P.A.: A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum. Factors **58**, 377–400 (2016)
5. Detection of functional brain network reconfiguration during task-driven cognitive states. Neuroimage. **142**, 198–210 (2016)
6. Garcia, J.O., Brooks, J., Kerick, S., Johnson, T., Mullen, T.R., Vettel, J.M.: Estimating direction in brain-behavior interactions: Proactive and reactive brain states in driving. NeuroImage **150**, 239–249 (2017)
7. Kalia, A.K., Buchler, N., DeCostanza, A., Singh, M.P.: Computing Team Process Measures From the Structure and Content of Broadcast Collaborative Communications. IEEE Trans. Comput. Soc. Syst. **4**, 26–39 (2017)
8. Bonato, P.: Wearable sensors and systems. IEEE Eng. Med. Biol. Mag. **29**, 25–36 (2010)
9. Chen, J.Y., Haas, E.C., Barnes, M.J.: Human performance issues and user interface design for teleoperated robots. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**, 1231–1245 (2007)
10. Dean, R.M.S.: Common world model for unmanned systems. In: Unmanned Systems Technology XV, p. 87410O. International Society for Optics and Photonics (2013)
11. Dean, R.M.S., Oh, J., Vinokurov, J.: Common world model for unmanned systems: phase 2. In: Unmanned Systems Technology XVI, p. 90840I. International Society for Optics and Photonics (2014)
12. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors J. Hum. Factors Ergon. Soc. **46**, 50–80 (2004)
13. de Visser, E.J., Cohen, M., Freedy, A., Parasuraman, R.: A design methodology for trust cue calibration in cognitive agents. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 251–262. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_24
14. Bitan, Y., Meyer, J.: Self-initiated and respondent actions in a simulated control task. Ergonomics **50**, 763–788 (2007)
15. Seppelt, B.D., Lee, J.D.: Making adaptive cruise control (ACC) limits visible. Int. J. Hum.-Comput. Stud. **65**, 192–205 (2007)
16. Stanton, N.A., Walker, G.H., Young, M.S., Kazi, T., Salmon, P.M.: Changing drivers' minds: the evaluation of an advanced driver coaching system. Ergonomics **50**, 1209–1234 (2007)
17. Linegang, M.P., Stoner, H.A., Patterson, M.J., Seppelt, B.D., Hoffman, J.D., Crittendon, Z.B., Lee, J.D.: Human-automation collaboration in dynamic mission planning: a challenge requiring an ecological approach. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 2482–2486. SAGE Publications, Los Angeles (2006)
18. Chen, J.Y., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.: Situation awareness-based agent transparency. Army Research Laboratory (2014)
19. Chen, J.: Advances in Human Factors in Robots and Unmanned Systems (2018)
20. Schaefer, K.: The perception and measurement of human-robot trust (2013)
21. Schaefer, K.E., Straub, E.R.: Will passengers trust driverless vehicles? Removing the steering wheel and pedals. In: 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), pp. 159–165. IEEE (2016)
22. Kostavelis, I., Gasteratos, A.: Semantic mapping for mobile robotics tasks: a survey. Robot. Auton. Syst. **66**, 86–103 (2015)
23. Wolf, D.F., Sukhatme, G.S.: Semantic mapping using mobile robots. IEEE Trans. Robot. **24**, 245–258 (2008)

24. Heitz, G., Koller, D.: Learning spatial context: using stuff to find things. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 30–43. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88682-2_4
25. Meger, D., Forssén, P.-E., Lai, K., Helmer, S., McCann, S., Southey, T., Baumann, M., Little, J.J., Lowe, D.G.: Curious george: an attentive semantic robot. Robot. Auton. Syst. **56**, 503–511 (2008)
26. Walter, M.R., Hemachandra, S., Homberg, B., Tellex, S., Teller, S.: Learning semantic maps from natural language descriptions (2013)
27. Maniadakis, M., Trahanias, P.: Time models and cognitive processes: a review. Front. Neurorobot. **8**, (2014)
28. Lee, J.: Trust, trustworthiness, and trustability. In: Presentation at the Workshop on Human Machine Trust for Robust Autonomous Systems (2012)
29. Cook, M.B., Smallman, H.S.: Human factors of the confirmation bias in intelligence analysis: decision support from graphical evidence landscapes. Hum. Factors **50**, 745–754 (2008)
30. Neyedli, H.F., Hollands, J.G., Jamieson, G.A.: Beyond identity: incorporating system reliability information into an automated combat identification system. Hum. Factors **53**, 338–355 (2011)
31. Tufte, E.R.: The Visual Display of Quantitative Information. Graphics Press, Cheshire (2001)
32. Schaefer, K.E.: Measuring trust in human robot interactions: development of the "*Trust Perception Scale-HRI*". In: Mittu, R., Sofge, D., Wagner, A., Lawless, W.F. (eds.) Robust Intelligence and Trust in Autonomous Systems, pp. 191–218. Springer, Boston, MA (2016). https://doi.org/10.1007/978-1-4899-7668-0_10
33. Drnec, K., Metcalfe, J.S.: Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. In: Schmorrow, D.D.D., Fidopiastis, C.M.M. (eds.) AC 2016. LNCS (LNAI), vol. 9744, pp. 157–167. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39952-2_16
34. Gremillion, G.M., Metcalfe, J.S., Marathe, A.R., Paul, V.J., Christensen, J., Drnec, K., Haynes, B., Atwater, C.: Analysis of trust in autonomy for convoy operations. In: SPIE Defense + Security, p. 98361Z. International Society for Optics and Photonics (2016)
35. Nothwang, W.D., Gremillion, G.M., Donavanik, D., Haynes, B., Atwater, C., Canady, J., Metcalfe, J.S., Marathe, A.R.: Multi-sensor fusion architecture for human-autonomy teaming. In: 2016 9th International Symposium on Resilient Control Systems (ISRCS). IEEE (2016)
36. Schaefer, K.E., Straub, E.R., Chen, J.Y., Putney, J., Evans, A.: Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cogn. Syst. Res. **46**, 26–39 (2017)
37. Metcalfe, J., Marathe, A., Haynes, B., Paul, V., Gremillion, G., Drnec, K., Atwater, C., Estepp, J., Lukos, J., Carter, E., et al.: Building a framework to manage trust in automation. In: Micro- and Nanotechnology Sensors, Systems, and Applications IX, p. 101941U. International Society for Optics and Photonics (2017)
38. Perelman, B., Evans III, A., Schaefer, K.: Mental model consensus and shifts during navigation system-assisted route planning. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1183–1187. SAGE Publications, Los Angeles (2017)
39. Fitts, P.M.: Human engineering for an effective air-navigation and traffic-control system (1951)
40. Sheridan, T.B.: Function allocation: algorithm, alchemy or apostasy? Int. J. Hum.-Comput. Stud. **52**, 203–216 (2000)
41. Cummings, M.M.: Man versus machine or man + machine? IEEE Intell. Syst. **29**, 62–69 (2014)

42. Billings, C.E.: Human-centered aircraft automation: a concept and guidelines (1991)
43. Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. MIT Press, Cambridge (1992)
44. Fong, T., Thorpe, C., Baur, C.: Multi-robot remote driving with collaborative control. IEEE Trans. Ind. Electron. **50**, 699–704 (2003)
45. Abbink, D.A., Mulder, M., Boer, E.R.: Haptic shared control: smoothly shifting control authority? Cogn. Technol. Work **14**, 19–28 (2012)
46. Woods, D.D., Branlat, M.: Hollnagel's test: being "in control" of highly interdependent multi-layered networked systems. Cogn. Technol. Work **12**, 95–101 (2010)
47. Schaefer, K.E., Brewer, R.W., Putney, J., Mottern, E., Barghout, J., Straub, E.R.: Relinquishing manual control: collaboration requires the capability to understand robot intent. In: 2016 International Conference on Collaboration Technologies and Systems (CTS), pp. 359–366. IEEE (2016)
48. Woods, D.D.: Cognitive technologies: the design of joint human-machine cognitive systems. AI Mag. **6**, 86 (1985)
49. Cummings, M., Clare, A.: Holistic modelling for human-autonomous system interaction. Theor. Issues Ergon. Sci. **16**, 214–231 (2015)
50. Marathe, A.R., Metcalfe, J.S., Lance, B.J., Lukos, J.R., Jangraw, D., Lai, K.T., Touryan, J., Stump, E., Sadler, B.M., Nothwang, W., McDowell, K.: The privileged sensing framework: A principled approach to improved human-autonomy integration. Theor. Issues Ergon. Sci. **19**(3), 283–320 (2018)
51. Billings, C.E.: Aviation automation: The search for a human-centered approach. Lawrence Erlbaum Associates Inc., Mahwah (1997)
52. Dekker, S.W., Woods, D.D.: MABA-MABA or abracadabra? Progress on human–automation co-ordination. Cogn. Technol. Work **4**, 240–244 (2002)
53. Lyons, J.B.: Being Transparent about Transparency (2013)
54. Chen, J.Y., Barnes, M.J.: Human–agent teaming for multirobot control: a review of human factors issues. IEEE Trans. Hum.-Mach. Syst. **44**, 13–29 (2014)
55. Canady, J.D., Marathe, A.R., Herman, D.H., Files, B.T.: A maximum likelihood method for estimating performance in a rapid serial visual presentation target-detection task. In: Schmorrow, D., Fidopiastis, C.M. (eds.) VAMR 2018. LNCS 10909, pp. 383–392. Springer, Heidelberg (2018)
56. Schaefer, K.E., Perelman, B.S., Brewer, R.W., Wright, J., Roy, N., Aksaray, D.: Quantifying human decision-making: implications for bidirectional communication in human-robot teams. In: Schmorrow, D., Fidopiastis, C.M. (eds.) VAMR 2018. LNCS 10909, pp. 361–379. Springer, Heidelberg (2018)
57. Drnec, K., Gremillion, G.M., Donavanik, D., Canady, J.D., Atwater, C.S., Haynes, B.A., Marathe, A.R., Carter, E.C., Metcalfe, J.S.: The role of psychophysiological measures as implicit communication within mixed-initiative teams. In: Schmorrow, D., Fidopiastis, C.M. (eds.) VAMR 2018. LNCS 10909, pp. 299–313. Springer, Heidelberg (2018)