Situational Adaptive Autonomy in Human-AI Cooperation

Vildan Salikutluk[1,2,3], Eric Frodl[1], Franziska Herbert[1], Dirk Balfanz[1,2] and Dorothea Koert[1,2]

[1]Interactive AI Algorithms & Cognitive Models for Human-AI Interaction, Computer Science Department, Technical University Darmstadt
[2]Models of Higher Cognition, Human Sciences Department, Technical University Darmstadt
[3]Centre for Cognitive Science, Technical University Darmstadt

Abstract
Human-AI teams have the potential to produce improved outcomes in various tasks as opposed to each team member working alone. However, there are various factors that influence human-AI team performance which potentially differ from classical Human-Computer Interaction settings. Specifically, there is existing work indicating that it is beneficial for AI systems to automatically adapt their autonomy within the team and task in order to work towards achieving a shared goal more effectively. Thus, in this paper, we describe a concept of situational adaptive autonomy for human-AI cooperation in a shared workspace setting. We discuss that task complexity and models for the AI system's understanding of their human teammate, i.e. Theory of Mind models (ToMMs) might be helpful to implement situational adaptation of AI autonomy such that interaction and team performance can be improved. We present an experimental setup for a cooperative real-world robotic task and a corresponding approximation in a grid-world in which we plan to investigate situational adaptive autonomy within a shared workspace in an interactive human-AI team.

Keywords
Human-AI Cooperation, Interactive Human-AI Teaming, Adaptive Autonomy, Collaborative Problem-Solving, Robotic Task

1. Introduction

Recent artificial intelligence (AI) systems and robots as their embodied form show great potential to support humans at work [1, 2, 3] or in everyday life tasks [4, 2, 5, 6, 7, 8]. However, improving task outcomes through human-AI collaboration is not trivial, often case-specific, and depends on the abilities and characteristics of each party [9]. While general design principles for classical human-computer interaction (HCI) have already been explored well, there is a need to update them for designing interactions with AI systems [10, 11, 12] such that human-AI teams can solve problems synergistically and improve their performance together. Hereby, it is important to consider factors where current and future AI systems might differ from classical HCI systems.

One distinguishing factor is the higher degree of autonomy that AI systems can possibly achieve during cooperative tasks [13]. When humans interact with technical systems as tools, they commonly automate a very specific sub-task to achieve their overall goals more efficiently by using the system [14, 15, 16]. In such cases, systems have a specific and limited purpose but in general no autonomy within the task and the team. On the other end of the autonomy spectrum lies a fully autonomous collaboration partner with whom the human outputs a collective result for a shared goal, i.e. human-human interaction, e.g. when colleagues collaborate with each other. Such collaboration depends on communication, the skill level of each partner, etc. [17, 18]. Often, the interaction with AI agents falls somewhere in between the ends of the spectrum shown in Fig. 1.

Different concepts for autonomy levels have also been proposed in previous work [19, 13, 20, 21]. However, high(er) autonomy in systems does not necessarily increase team performance or is preferred by their human counterparts in every situation [22]. Previous work also indicated that the ability to slide along the autonomy scale and dynamically adapt autonomy levels is beneficial [23, 21, 13, 24]. In more recent work, autonomy is often specified as a set of autonomy levels which an operator can switch (manually) [19, 20, 21]. Automatically adjusting autonomy already showed promising results in specific use cases, e.g. multi-agent systems without human interaction [23], Unmanned Aerial Vehicle path-planning [21], settings where an operator remotely controls a robot in hazardous environments [13] or simulation-based evaluations for a cleaning and an inventory scenario with a mobile manipulator [24]. However, we see a lack of experimental real-world evaluations of concepts for situational adaptation of AI autonomy in cooperative shared workspace settings.

Therefore, with our planned experiments we aim to contribute empirically validated guidelines for situational adaptation of autonomy in shared workspaces that facilitate improved team performance and interaction within the human-AI team. In particular, we suggest to incorporate measures for task complexity as well as models for the AI system's understanding of their human teammate,

AutomationXP23: Intervening, Teaming, Delegating - Creating Engaging Automation Experiences, April 23rd, Hamburg, Germany
This work was funded by German Federal Ministry of Education and Research (project IKIDA, 01IS20045).
vildan.salikutluk@tu-darmstadt.de (V. Salikutluk)
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)
CEUR Workshop Proceedings (CEUR-WS.org)

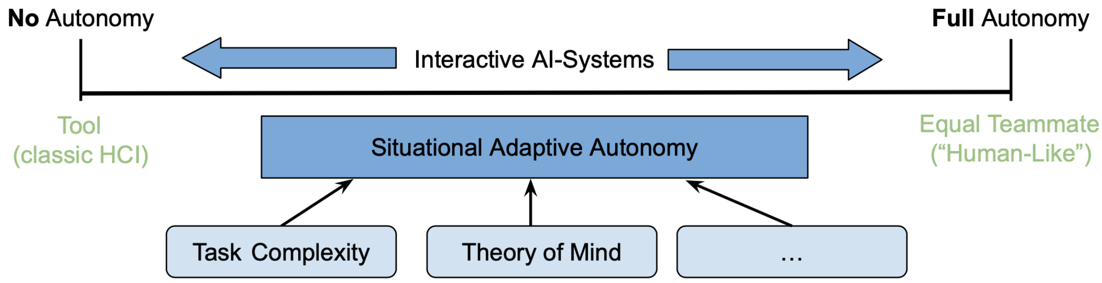


Figure 1: Overview of sliding degree of AI autonomy. It can range from no autonomy (left), i.e. like a tool completely controlled by its user to fully autonomous (right) where a technical system becomes a somewhat equal teammate for the human. It is important to investigate the interaction paradigms arising when interactive AI systems slide along this autonomy scale. We propose to investigate the situational adaptation of autonomy in cooperative shared workspace settings depending on factors such as the overall task complexity and the AI system's understanding of its human partner (Theory of Mind models).

i.e. Theory of Mind models [25, 26, 27], to automatically decide when the AI system should switch its autonomy level in collaborative interaction with their human partner. This should ensure that interaction is initiated and expressed by the AI system only when it is appropriate and helpful for the human and the overall task goal. The proposed concept for situational adaptive autonomy results in an interactive AI system which can slide on the autonomy spectrum shown in Fig. 1 in blue.

We advocate for investigating interaction paradigms arising when interactive AI systems slide along this autonomy scale. To this end, we propose an experimental task setup within a grid-world environment and a real-world robotic task to gain deeper insights into the interaction with a coordinating human-AI team.

2. Background

There is a large body of previous work defining autonomy and its possible levels for AI or robotic agents [28, 22, 29, 30, 31, 32, 33]. For a more detailed overview on the definition of autonomy that aligns well with how it is framed in our work we refer the interested reader to [31, 32]. In this section, we focus on the discussion of related work that aims to enable AI agents to automatically adjust their level of autonomy (Section 2.1) and subsequently provide a short background on *task complexity* and *Theory of Mind models*, which we propose as two important factors to implement situational adjustment of autonomy in shared workspace settings (Section 2.2).

2.1. Adaptive Autonomy

Dynamic adjustment of autonomy has been explored a lot in settings where multiple AI agents collaborated [23]. Additionally, automatic adjustment of an AI agent's autonomy has been shown to be beneficial in settings where a human operator remotely controls one or multiple robots in hazardous or space environments [34, 34] and for path planning of Unmanned Aerial Vehicles [21]. In particular, [21] shows that human-AI cooperation with autonomy adaptation can lead to better performance as opposed to either human or system completing the task alone.

Generally, there is some evidence that humans profit from [35] and are in favor of systems having a high(er) level of autonomy [13] when it helps their goals. Humans also share their task load more when they perceive a system's behavior as human-like [36]. Nevertheless, there is also literature about how humans sometimes prefer when they have control over systems [37, 10] or reduce the systems' autonomy [35]. Also, the interplay of a system's autonomy level and the situational awareness of its user has been investigated [38]. Further work also acknowledges that the design and evaluation of human-AI interaction (partly) depends on the autonomy spectrum on which systems can lay from just being tools to being counterparts or teammates [39].

Some research also demonstrates the potential of adaptive autonomy in shared workspace settings [40, 24]. In [40], the authors present a framework that incorporates situation assessment and planning together with human intention prediction and reactive action execution. Their approach enables a robot to adapt to user preferences allowing the human partner to be more passive or active in giving commands. A Theory of Mind model for predicting temporary absence or inattention of the human is proposed in [24] to automatically adapt robot communication patterns during the execution of a cooperative table cleaning task. However, both approaches are only evaluated in virtual environments with simulated humans. We found a lack of evaluations for adjustable autonomy in shared workspaces with real human users and in real-world robotic scenarios.

2.2. Task Complexity & Theory of Mind

There are various existing definitions for task complexity [41, 42, 43, 44]. The definitions often aim to describe how much cognitive processing capabilities, skills, information, and knowledge availability are necessary to perform a task [45, 41]. Furthermore, objective task complexity can comprise of number of task components, goals, or possible solution paths [46, 47]. Overall, [48] consolidate existing definitions into one model for task complexity. It comprises of ten dimensions, e.g. the number of task components, their interdependency and time-related constraints. Furthermore, there are frameworks for task complexity for human-system integration [49] and for human teams [47]. Specifically, [47] distinguishes between component complexity which considers task aspects for the individuals within the team and coordinative complexity regarding factors of interaction and teamwork such as interdependencies and solution diversity.

Task complexity influences individual and group performance in different tasks [48, 50]. While teams perform worse compared to individuals when working on lower complexity tasks, higher task complexity leads to interacting groups outperforming individual performance [50]. This means, humans can profit from support when working on tasks with higher task complexity. This is also in line with [51] who found that humans rely more on support of systems when there is higher task complexity. While there are also other aspects which influence the performance of human-AI teams [52], they might be most successful when humans and AI agents complement each other [53, 54]. Complementarity requires among other things that teammates have awareness of the situation [55, 56] and about what their teammate knows and plans, which is known as Theory of Mind, i.e. the modeling of mental states of others [25]. They are also used computationally in various human-AI interaction settings [26, 27, 24, 57, 58] such that they allow for anticipation for human actions and planning with an appropriate model of the human in the interaction. This can ensure that systems can adjust for specific users and plan better (together) with them [24].

3. Situational Adaptive Autonomy for Cooperative Tasks in Shared Workspaces

Situational adaptive autonomy may influence the interaction and performance of human-AI teams in cooperative tasks. In this section, we discuss how we plan to realize a concept of situational adaptive autonomy in a shared workspace setting (Section 3.1) and describe the task in which we plan to evaluate our approach (Section 3.2).

3.1. Initiative and Delegation within the Human-AI Team

For our approach and all planned experiments, there is no debating or adjusting of the goal itself, i.e. the human sets the overall task goal, which is to be achieved with the help of the AI teammate [19] and the goal is known to both. In order to investigate effects on the interaction paradigms for situational adaptive autonomy in a shared workspace, we first of all require a concrete implementation of autonomy levels for the AI agent. Existing concepts for autonomy levels [22, 38, 59, 60, 32, 28, 29] have often been proposed in a more theoretical framing. All of these concepts include at least 10 levels of autonomy for an AI agent.

For our concrete implementation we decided to follow the concept from [29, 28] but summarize their 10 levels into four. This is due to two reasons: First, while [29, 28] distinguish some levels by how and if the robot informs the human about its choice, we implement the AI system's decision on how and when to notify or question the human based on a ToMM. This model considers the situation and current state of its human partner and thus only informs or asks them about anything when it assumes that the human is able to comprehend and reply. Thus, we do not separate such communication with the human partner in distinct autonomy levels. Second, we hypothesize that it is easier for the human teammate to understand in which autonomy level the AI partner currently is when there are less levels overall.

We implement the expression of the AI system's autonomy level by how much initiative it takes to deviate from its currently assigned sub-task if it assumes a benefit for the team performance. We specify the following four different types of initiative.

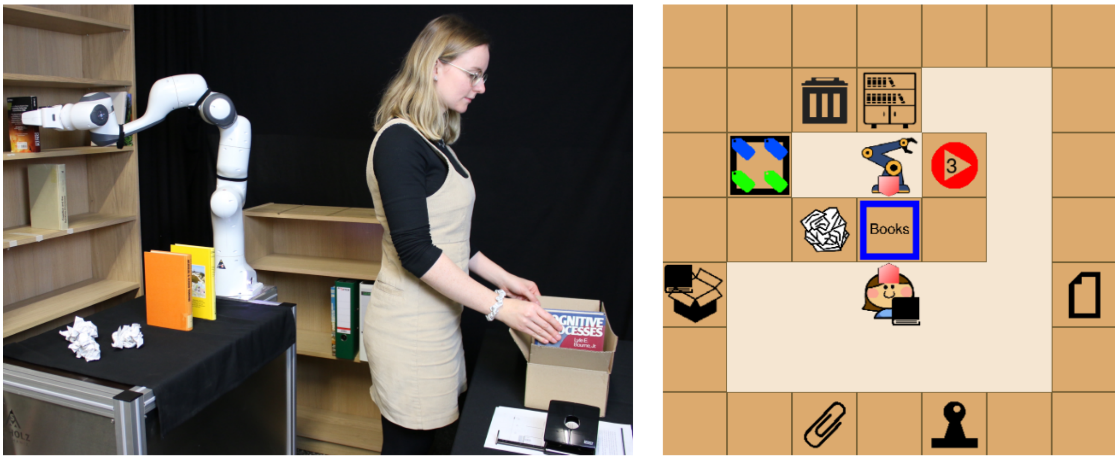
No Initiative: the AI just continues with its current sub-task; makes no suggestions if it notices sub-tasks with higher priority; goes into idle mode if it encounters a problem during its sub task execution.

Low Initiative: If during a sub-task execution the AI encounters a problem or notices a sub-task with a higher priority, it presents a list of possible alternative sub-tasks to the human and waits to see if they choose one.

Moderate Initiative: If during a sub-task execution the AI encounters a problem or notices a sub-task with a higher priority it presents the option it assumes as the best possibility and waits for confirmation (or rejection).

Full Initiative: If during a sub-task execution the AI encounters a problem or notices a sub-task with a higher priority, the AI executes an alternative it considers the best possibility for improvement; if the ToMM indicates human availability it informs the human and ask them if the decision was okay.

Overall, assigning sub-tasks can be based on access, competence and permission to execute them as also pro-



(a) (b)

Figure 2: Overview of our experimental setup for investigating effects of situational autonomy adaptation in cooperative tasks within shared workspaces. (a) Real-world robotic task in the lab. (b) Approximated task setup in a grid-world setting. The human is depicted by the blue agent. The AI is visualized as an robotic arm.

posed in [31]. The AI system can lower its autonomy when uncertainties or problems arise or if a miss match between its own competence and (sub-)task requirements occurs. Additionally, humans can intervene in the AI's actions or potentially delegate new sub-tasks. An increase in autonomy can be beneficial, e.g. when only execution of a sub-task with higher priority can prevent catastrophic failure of the overall task and the ToMM indicates current unavailability or missing situational awareness of the human. We theorize that when task complexity is low(er), autonomy and initiative can be low(er) as well while the team could profit from high(er) autonomy and initiative from the AI when task complexity is high(er), as in [50].

3.2. Experimental Task Setting

We consider a setting in a shared workspace where the overall task consists of sub-tasks that can be either performed by only the AI agent, only the human or both but in some cases with potentially different degree of efficiency. Additionally, we assume that there is always a set of sub-tasks that may result in or prevent failure of the overall task goal. Our setups for the robotic task and the corresponding abstracted grid-world, inspired by [61], are illustrated in Fig. 2. In the proposed task, the human and AI need to organize and process the contents of boxes which get delivered over time. These boxes contain various objects, such as books or documents that require individual handling before they can be stored in their designated spaces. Task performance can be measured e.g. by the number of completely sorted boxes over a predefined time. This task cannot only vary situationally in its complexity but also requires ToMM for successful collaboration.

4. Implications and Future Work

For the development of interaction paradigms and a more common definition of situational autonomy adaptation in shared workspace settings, we consider it crucial to gain empirical insights from evaluations with real humans in real robotic tasks. To this end, we plan to implement the concept described in this paper and evaluate it with the discussed task setting. We strive to gain a deeper understanding of how factors such as task complexity or ToMM can possibly be used for situational autonomy adaptation in shared workspaces. Particularly, we consider it important to investigate the effects of the resulting situational autonomy adaptation on the performance of and interaction within human-AI teams. We plan to test whether our conceptualized and implemented autonomy levels and corresponding initiative types positively impact team performance in human-AI interaction for a collaborative task within a shared workspace.

Generally, we advocate that the successful deployment of situational autonomy adaptation in human-AI interaction requires more interdisciplinary exchange and research to better understand the implications for both humans and AI systems in the future.

References

[1] N. Anantrasirichai, D. Bull, Artificial intelligence in the creative industries: a review, Artificial intelligence review (2022) 1–68.

[2] A. Maedche, C. Legner, A. Benlian, B. Berger, H. Gimpel, T. Hess, O. Hinz, S. Morana, M. Söllner, Ai-based digital assistants: Opportunities, threats, and research perspectives, Business & Information Systems Engineering 61 (2019) 535–544.

[3] P. Mirowski, K. W. Mathewson, J. Pittman, R. Evans, Co-writing screenplays and theatre scripts with language models: An evaluation by industry professionals, arXiv preprint arXiv:2209.14958 (2022).

[4] K. Yamazaki, R. Ueda, S. Nozawa, M. Kojima, K. Okada, K. Matsumoto, M. Ishikawa, I. Shimoyama, M. Inaba, Home-assistant robot for an aging society, Proceedings of the IEEE 100 (2012) 2429–2441.

[5] X. Guo, Z. Shen, Y. Zhang, T. Wu, Review on the application of artificial intelligence in smart homes, Smart Cities 2 (2019) 402–420.

[6] G. Terzopoulos, M. Satratzemi, Voice assistants and smart speakers in everyday life and in education, Informatics in Education 19 (2020) 473–490.

[7] A. Hepp, Artificial companions, social bots and work bots: communicative robots as research objects of media and communication studies, Media, Culture & Society 42 (2020) 1410–1426.

[8] R. Macrorie, S. Marvin, A. While, Robotics and automation in the city: a research agenda, Urban Geography 42 (2021) 197–217.

[9] A. Campero, M. Vaccaro, J. Song, H. Wen, A. Almaatouq, T. W. Malone, A test for evaluating performance in human-computer systems, arXiv preprint arXiv:2206.12390 (2022).

[10] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, S. I. Jina Suh, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, E. Horvitz, Guidelines for human-ai interaction, in: CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), ACM, New York, NY, 2019. doi:10.1145/3290605.3300233.

[11] W. Xu, Toward human-centered ai: a perspective from human-computer interaction, Interactions 26 (2019) 42–46.

[12] W. Xu, M. J. Dainoff, L. Ge, Z. Gao, Transitioning to human interaction with ai systems: New challenges and opportunities for hci professionals to enable human-centered ai, 2021. URL: https://arxiv.org/abs/2105.05424. arXiv:2105.05424.

[13] P. Schermerhorn, M. Scheutz, Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task, in: Proceedings of the 2009 international conference on multimodal interfaces, 2009, pp. 63–70.

[14] S. K. Card, T. P. Moran, A. Newell, The psychology of human-computer interaction, Crc Press, 2018.

[15] D. Kieras, Goms models for task analysis, The handbook of task analysis for human-computer interaction (2004) 83–116.

[16] A. Ramkumar, P. J. Stappers, W. J. Niessen, S. Adebahr, T. Schimek-Jasch, U. Nestle, Y. Song, Using goms and nasa-tlx to evaluate human–computer interaction process in interactive segmentation, International Journal of Human–Computer Interaction 33 (2017) 123–134.

[17] B. Bahrami, K. Olsen, P. E. Latham, A. Roepstorff, G. Rees, C. D. Frith, Optimally interacting minds, Science 329 (2010) 1081–1085.

[18] D. Owen, Collaborative decision making, Decision Analysis 12 (2015) 29–45.

[19] V. Z. Moffitt, J. L. Franke, M. Lomas, Mixed-initiative adjustable autonomy in multi-vehicle operations, Proceedings of AUVSI, Orlando, Florida (2006).

[20] S. Zieba, P. Polet, F. Vanderhaegen, S. Debernard, Principles of adjustable autonomy: a framework for resilient human–machine cooperation, Cognition, Technology & Work 12 (2010) 193–203.

[21] L. Lin, M. A. Goodrich, Sliding autonomy for uav path-planning: adding new dimensions to autonomy management, in: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, 2015, pp. 1615–1624.

[22] M. R. Endsley, From here to autonomy: lessons learned from human–automation research, Human factors 59 (2017) 5–27.

[23] K. Suzanne Barber, A. Goel, C. E. Martin, Dynamic adaptive autonomy in multi-agent systems, Journal of Experimental & Theoretical Artificial Intelligence 12 (2000) 129–147.

[24] S. Devin, R. Alami, An implemented theory of mind to improve human-robot shared plans execution, in: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2016, pp. 319–326.

[25] H. M. Wellman, The child's theory of mind., The MIT Press, 1992.

[26] L. M. Hiatt, A. M. Harrison, J. G. Trafton, Accommodating human variability in human-robot teams through theory of mind, in: Twenty-Second International Joint Conference on Artificial Intelligence, 2011.

[27] J. M. Beer, A. D. Fisk, W. A. Rogers, Toward a framework for levels of robot autonomy in human-robot interaction, Journal of human-robot interaction 3 (2014) 74.

[28] T. O'Neill, N. McNeese, A. Barron, B. Schelble, Human–autonomy teaming: A review and anal-

ysis of the empirical literature, Human factors 64 (2022) 904–938.
[29] R. Parasuraman, T. B. Sheridan, C. D. Wickens, A model for types and levels of human interaction with automation, IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans 30 (2000) 286–297.
[30] H. A. Abbass, Social integration of artificial intelligence: Functions, automation allocation logic and human-autonomy trust, Cognitive Computation 11 (2019) 159–171.
[31] J. M. Bradshaw, P. J. Feltovich, H. Jung, S. Kulkarni, W. Taysom, A. Uszok, Dimensions of adjustable autonomy and mixed-initiative interaction, in: Agents and Computational Autonomy: Potential, Risks, and Solutions 1, Springer, 2004, pp. 17–39.
[32] S. A. Mostafa, M. S. Ahmad, A. Mustapha, Adjustable autonomy: a systematic literature review, Artificial Intelligence Review 51 (2019) 149–186.
[33] C. Castelfranchi, Founding agents'" autonomy" on dependence theory, in: ECAI, volume 1, 2000, pp. 353–357.
[34] G. Dorais, R. P. Bonasso, D. Kortenkamp, B. Pell, D. Schreckenghost, Adjustable autonomy for human-centered autonomous systems on mars, in: Mars society conference, 1998.
[35] S. S. Sundar, Rise of machine agency: A framework for studying the psychology of human–ai interaction (haii), Journal of Computer-Mediated Communication 25 (2020) 74–88.
[36] B. Wahn, A. Kingstone, Humans share task load with a computer partner if (they believe that) it acts human-like, Acta Psychologica 212 (2021) 103205.
[37] B. Y. Lim, A. K. Dey, Assessing demand for intelligibility in context-aware applications, in: Proceedings of the 11th international conference on Ubiquitous computing, 2009, pp. 195–204.
[38] M. R. Endsley, Automation and situation awareness, in: Automation and human performance: Theory and applications, CRC Press, 2018, pp. 163–181.
[39] C. Wienrich, M. E. Latoschik, extended artificial intelligence: New prospects of human-ai interaction research, Frontiers in Virtual Reality 2 (2021) 686783.
[40] M. Fiore, A. Clodic, R. Alami, On planning and task achievement modalities for human-robot collaboration, in: Experimental Robotics: The 14th International Symposium on Experimental Robotics, Springer, 2016, pp. 293–306.
[41] D. D. Woods, Coping with complexity: the psychology of human behaviour in complex systems, in: Tasks, errors, and mental models, 1988, pp. 128–148.
[42] N. R. Bailey, M. W. Scerbo, Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust, Theoretical Issues in Ergonomics Science 8 (2007) 321–348.
[43] F. L. Greitzer, Toward the development of cognitive task difficulty metrics to support intelligence analysis research, in: Fourth IEEE Conference on Cognitive Informatics, 2005.(ICCI 2005)., IEEE, 2005, pp. 315–320.
[44] P. Ø. Braarud, B. Kirwan, Task complexity: what challenges the crew and how do they cope, in: Simulator-based human factors studies across 25 years: The history of the halden man-machine laboratory, Springer, 2011, pp. 233–251.
[45] D. J. Campbell, Task complexity: A review and analysis, Academy of management review 13 (1988) 40–52.
[46] T. M. Brown, C. E. Miller, Communication networks in task-performing groups: Effects of task complexity, time pressure, and interpersonal dominance, Small group research 31 (2000) 131–157.
[47] E. H. Lazzara, D. Pavlas, S. Fiore, E. Salas, A framework to develop task complexity, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 54, SAGE Publications Sage CA: Los Angeles, CA, 2010, pp. 2338–2342.
[48] P. Liu, Z. Li, Task complexity: A review and conceptualization framework, International Journal of Industrial Ergonomics 42 (2012) 553–568.
[49] M. Efatmaneshnik, H. A. Handley, Task complexity measurement framework for human systems integration, IEEE Systems Journal 15 (2020) 2787–2797.
[50] A. Almaatouq, M. Alsobay, M. Yin, D. J. Watts, Task complexity moderates group synergy, Proceedings of the National Academy of Sciences 118 (2021) e2101062118.
[51] R. Parasuraman, V. Riley, Humans and automation: Use, misuse, disuse, abuse, Human factors 39 (1997) 230–253.
[52] O. Vereschak, G. Bailly, B. Caramiaux, How to evaluate trust in ai-assisted decision making? a survey of empirical methodologies, Proceedings of the ACM on Human-Computer Interaction 5 (2021) 1–39.
[53] M. Steyvers, H. Tejeda, G. Kerrigan, P. Smyth, Bayesian modeling of human–ai complementarity, Proceedings of the National Academy of Sciences 119 (2022) e2111547119.
[54] K. Holstein, V. Aleven, Designing for human-ai complementarity in k-12 education, 2021. URL: https://arxiv.org/abs/2104.01266. arXiv:2104.01266.
[55] M. Demir, N. J. McNeese, N. J. Cooke, Team situation awareness within the context of human-autonomy teaming, Cognitive Systems Research 46 (2017) 3–12.
[56] J. Jiang, A. J. Karran, C. K. Coursaris, P.-M. Léger, J. Beringer, A situation awareness perspective on

human-ai interaction: Tensions and opportunities, International Journal of Human–Computer Interaction (2022) 1–18.

[57] M. M. Çelikok, T. Peltola, P. Daee, S. Kaski, Interactive ai with a theory of mind, arXiv preprint arXiv:1912.05284 (2019).

[58] K. I. Gero, Z. Ashktorab, C. Dugan, Q. Pan, J. Johnson, W. Geyer, M. Ruiz, S. Miller, D. R. Millen, M. Campbell, et al., Mental models of ai agents in a cooperative game setting, in: Proceedings of the 2020 chi conference on human factors in computing systems, 2020, pp. 1–12.

[59] T. B. Sheridan, W. L. Verplank, Human and computer control of undersea teleoperators, Technical Report, Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, 1978.

[60] T. B. Sheridan, Telerobotics, automation, and human supervisory control, MIT press, 1992.

[61] S. A. Wu, R. E. Wang, J. A. Evans, J. B. Tenenbaum, D. C. Parkes, M. Kleiman-Weiner, Too many cooks: Bayesian inference for coordinating multi-agent collaboration, Topics in Cognitive Science 13 (2021) 414–432.