REQUIREMENTS FOR COMPUTATIONAL APPROACHES TO ANALYZING RESILIENCE IN HUMAN-MACHINE TEAMS

<author names and affiliation omitted for blind review>

Robotic and automated technologies are becoming increasingly recognized as teammates to human agents performing joint activities. Humans working alongside limited robotic/AI teammates changes the dynamics of work, introducing complexities and pitfalls that require new methods to evaluate the performance of the system. Human-machine teams (HMTs) should be recognized as distributed cognitive work systems in order to highlight the importance of interactions within work and between agents, which have been noted to be crucial predictors of team performance in homogenous human teams (Banks & Millward, 2009; Cooke et al., 2007, 2008). This paper outlines requirements and perspectives for novel modeling and simulation approaches to evaluate HMTs, drawing concepts from literature in human-robot-interaction (HRI), cognitive systems engineering, and work-centered computational models. We discuss the potential for operationalizing key aspects of resilience to support both evaluation of HMTs and the discovery of alternative architectural characteristics and governance protocols to increase sustained adaptability.

INTRODUCTION

Human-machine team teaming posits that novel forms of technology should be viewed from the perspective of a team member rather than a prosthesis or tool for extending human capabilities. However, potential pitfalls and complexities associated with limited (and sometimes brittle) teammates introduces new design challenges and underscores the need to view teams as distributed cognitive systems. Computational modeling and simulation can provide new approaches to the evaluation of human-machine teams performing complex joint activities, formalizing representations of work to support large-scale, quantitative evaluation.

This paper proposes requirements for developing computational modeling and simulation frameworks for evaluating HMTs, specifically focusing on understanding resilience in distributed cognitive systems. Resilience is a key aspect of HMT performance in complex domains. We conducted a literature survey of resilience engineering concepts and existing testbeds for multi-human, multi-machine interaction. We discuss the necessity for a novel and interdisciplinary approach to understand resilient performance in multi-agent sociotechnical systems such as HMTs.

BACKGROUND

Human-machine teams in complex work domains (such as aviation, healthcare, emergency response, space operations, and, to a lesser degree, manufacturing) need to be able to adapt to variable and uncertain work demands. The development of systems that can extend their performance beyond traditional performance boundaries is a key goal of the emerging field of resilience engineering (Woods, 2019). Resilience engineering is focused on preparing systems for critical failures and providing practitioners the tools to handle novel challenges as they arise. As such, the approach is explicitly focused on understanding how architecture and governance in distributed cognitive work systems affect adaptation, recognizing that performance of distributed work systems is emergent, including how weaknesses coalesce in challenging scenarios to cause cascading failures (Woods, 2015).

Several definitions of resilience have emerged in literature, though the two manifestations most relevant to supporting adaptation in complex work domains are sustained adaptability and graceful extensibility of a particular work system and the agents within it (Woods, 2015). Sustained adaptability (or resilience as the opposite of brittleness) is interested in the various dimensions of human performance and the associated tradeoffs that are engendered when building a system. It relates to the ability to continuously adapt in the face of changing environmental demands. Graceful extensibility, on the other hand, refers to how our systems stretch to handle surprises at the edge of their performance boundaries (Woods, 2018). This paper will focus on the idea of sustained adaptability, as this concept for resilience is noted to relate closely to the interactions of agents within the system and their ability to coordinate and collaborate in real time (Woods, 2015).

The theoretical underpinnings of Woods' resilience hold a key assumption: that the robot/AI teammate is neither totally reactive nor predictable (Bainbridge, 1983; Feigh et al., 2007; Hollnagel & Woods, 1983). Since machines may not always exhibit observability, predictability, and directability, and thus have the potential to be ineffective teammates, engineering HMTs with adaptability in mind can afford practitioners greater support when facing novel and challenging work environments. Systems that support sustained adaptability are able to rapidly shift their strategies, for example through changing information exchange modes, the distribution of authority and responsibility, and more as required by the demands of the environment. As all systems are resource limited, identifying tradeoffs associated with various system architectures and strengthening contextually appropriate responses can combat natural tendencies for complex sociotechnical systems to be brittle (Klein et al., 2004; Woods, 2015).

Though many existing testbeds have homed in on the need for designing HMTs for joint activity, few (if any) appear to have incorporated the concept of resilience as a crucial aspect of evaluating HMTs in joint activity. Existing testbeds in HMT span several work domains, with varying ecological validities, research interests, and simulated/actual robotic capabilities. One testbed by Gervits et al. (2017), for example, favors a low overhead, chat-bot based HMT task in a manned spaceflight setting, boasting a low overhead testbed capable of probing initial inquiry into coordinative and collaborative action in the Human-Robot Interaction (HRI) space. Other existing frameworks boast complex software architectures to support distributed teamwork (Gervits et al., 2020; Schermerhorn et al., 2007). Generally, HRI testbeds focus on the interaction of humans and robots in terms of trust in automation (N. Wang et al., 2015), usability testing and teleoperation capacities (Oishi et al., 2011; J. Wang et al., 2003), or focus on demands imposed by teammates of a new nature (Mingyue Ma et al., 2018).

Interestingly, a case can be made that a perspective focused on resilience and sustained adaptability would encompass most, if not all, of the aforementioned technical foci. A recurring pattern within these testbeds can be observed, as interactions within work and factors such as team coordinative and collaborative behavior, task delegation, etc., largely drive the focal point of each testbed, just at varying junctions and levels of abstraction or emergence. Trust in automation refers to a practitioner's mental model of their machine teammates competency and is therefore an emergent feature of how agents interact within work. A team that employs automation in scenarios where it is bound to exhibit brittleness will engender low trust, thus the focus should be shifted to be proactive rather than reactive.

Novel approaches actively identifying architectural characteristics and governance protocols leading to resilient behavior could prove a powerful way to evaluate the adaptive capacity of HMTs, and identify brittleness, complexities, and pitfalls introduced when working alongside limited teammates. Structured approaches that support researchers and developers in operationalizing the core principles of resilience is an important step towards applying lessons-learned in resilience engineering to the design of HMTs. Computational modeling is one approach to operationalize these principles, particularly as computational power could afford the ability to test system architectures in 'fast-time' simulations and support identifying interdependencies to bolster resilience and sustained adaptivity. These postulated benefits, among others, show promise for the future of computational modeling and simulation for HMT design and evaluation.

REQUIREMENTS FOR EVALUATING RESILIENCE IN JOINT ACTIVITY

We propose a set of requirements for developing computational modeling and simulation frameworks for HMTs, specifically incorporating resilience as a key aspect of successful joint activity. These requirements aim to bridge gaps associated with increasing reliance on robotic/AI capabilities, aiming to put forth frameworks to analyze the sustained adaptability of HMTs.

Requirement 1: A framework should focus on the distributed cognitive work system

Frameworks for HMT should recognize teaming as a complex sociotechnical phenomenon. This means understanding the human-machine team as a distributed cognitive work system (DCWS), defined by Woods and Hollnagel as "an adaptive system which functions using knowledge about itself and the environment in the planning and modification of actions" (Hollnagel & Woods, 1999). Conceptualizing HMTs as DCWS brings an interaction-centered perspective to HMT evaluation, positing that the configuration and organization of the broader work system is a crucial driver of its performance (Vicente, 1999). As such, evaluative frameworks should expand to encapsulate a work-centered representation of HMTs, mapping how humans and robots actively inform and support each other's work, adapt to changes in the environment, and modify activities to support tasks at hand.

A DCWS focus should also enable the evaluation of teams at different levels or echelons of work, allowing multi-faceted analysis that can understand the bigger picture as well as the minutiae of work. As such, any evaluation framework should not have a predefined notion of what the system is – both in terms of the specific composition of the system as well as the aggregation in terms of organizational echelons. Thus, an effective framework has flexibility in scaling the system from traditional dyad to multi-human, multi-machine teams, and also allows for aggregating agents into higher-echelon groups or teams (e.g., to consider interaction between teams as elements of a system). Such a flexible approach will promote an extendable framework that can be useful even when new technology or modus operandi are introduced.

Requirement 2: A framework should capture the interdependent nature of activity in DCWSs

A key insight of a DCWS view is recognizing the importance of interdependencies within work. Interdependence between agents is a requisite feature of joint activity, often determining the choreography of work (Hollnagel & Woods, 1999; Johnson, 2014). As agents work toward common goals, their individual goals, limitations, and constraints coalesce, resulting in a complex web of teamwork dynamics and dependencies. The incorporation of limited teammates (robots/AI) further underscores the need to identify these connections, as brittleness or unexpected edge cases that topple a limited agent may lead to cascading failures across the distributed team (Klein et al., 2004). Such concerns point directly to the necessity of engineering HMTs with resilience in mind, as modeling and simulation techniques hold enormous potential to identify broader patterns of emergent behaviors and brittleness that may lead to cascading failures in the face of novel conditions (Woods, 2015; Ma et al., 2018). Using models and simulations to study the interdependence in HMTs will provide valuable insight as well as the opportunity to preparing responses to cascading failures within work promoting the sustained adaptability of the HMT.

Interdependence also impacts a team's ability to absorb perturbations and change strategies. Several inter-team aspects influence adaptation capabilities, including initiative—willingness to adapt or seize opportunities—and reciprocity—as a commitment to mutual assistance, and the basic compact—an agreement (often tacit) to facilitate teamwork (Klein et al., 2005; Woods, 2019). Cooke et al. (2007, 2008) also explicitly highlight interactions between team members as crucial predictors of team performance, mentioning the dynamics of work as a promising approach to understanding joint activity (also see next requirement). Other concepts that tie into a perspective of interdependence is team cognition (Cooke et al., 2007; Salas & Fiore, 2004). Team cognition is often thought of in terms of distributed mental models, focusing on the means by which information is disseminated amongst teams performing joint activity (Banks & Millward, 2009).

Requirement 3: A framework should capture the temporal dynamics of HMTs

Work within HMTs is inherently complex. In order to adequately represent the complexities of the task domain, a perspective that can capture domain-specific intricacies of work is required. These intricacies, or dynamics, are a result of intimate relationships between resources, agents, task completion, communication, and other significant facets shaping the behavior of the DCWS. This interplay includes the temporal progression of system states, as sequential actions performed by agents modify the work environment (and then the agents themselves) in a dialectical fashion; each action precedes and informs the next (IJtsma, 2019). To most accurately examine DCWS, a temporal element is necessary to capture emergent interactions. Thus, examining the dynamics of DCWS should include temporal dynamics. This also promotes the ability to recognize broader pattern within work and assess their suitability for joint activity in HMT.

The dynamics of DCWS constitute an intricate web of connections including resource constraints, temporal execution of tasks, and many more aspects. The way the agents in the system interact, communicate, and execute are some examples. Modeling and simulation can and should explicitly account for dynamic (and emergent) effects when mapping the limits of various system architectures. Examples of relevant interdependencies that can lead to dynamic effects include relationships between agents due to resource constraints, physical/cognitive load for agents performing tasks, and information exchange speed, among others (Hollnagel & Woods, 2005; Johnson et al., 2014; Klein et al., 2004).

Requirement 4: A framework should support exploration as part of a formative process

An effective framework supports researchers and designers in gaining insight into the complexities of resilient behavior in HMTs. Vicente distinguished between two approaches to design: normative and formative (Vicente, 1999). A formative framework supports the design process through a process of exploration – supporting the identification of relevant constraints and characteristics of the work domain, the work itself, and the DCWS. The understanding gained from a formative approach can then be translated into design requirements. A formative approach contrasts with a normative approach that is aimed at prescribing one particular optimal solution to a designer (and, with it, to the practitioner/worker), without necessarily supporting the designer in improving their understanding of the system and context. The complexity of DCWS necessitates a formative approach over a normative approach, as normative frameworks lead to oversimplification of the complexities of actual work-in-context.

As a formative tool, a framework should support exploration (most likely through iteration) for discovery of potential system architectures and their subsequent evaluation. The fast-time evaluation capabilities afforded by computational modeling and simulation can be a real strength in the exploration of constraints, architectures, protocols, or operational assumptions. To be useful, however, a framework should be easily reconfigurable, allowing developers to explore the problem space by changing relevant parameters (for example, the system architecture, the constraints, or the demands imposed on the system). The framework should also afford flexibility in the output that a model produces, supporting easy reconfiguration in terms of metrics that are deemed relevant for the problem that is being explored.

A formative process that uses computational modeling will likely involve multiple other methods for exploration. In particular, Woods and Roth (1994) note that cognitive simulations must be part of a cycle of field research and model-based investigations. An effective framework should therefore support integration with data obtained from empirical field studies, both during model development (using empirical data as input) and as part of an investigative and iterative cycle in which insights derived from the model lead to new questions to be pursued in field research and vice versa (Woods & Roth, 1994).

Requirement 5: A framework should support evaluation at different levels of abstraction and aggregation

Related to the previous requirement, a framework must be scalable and configurable to test HMTs at various level of abstraction and aggregation. Levels of abstraction here denote the language used to describe a system, ranging from the highest-level purposes of the DCWS to its physical representation and form. An effective framework recognizes that resilient behavior involves parallel and interrelated processes of sustained adaptability at multiple levels of abstraction. For example, adaptability in HMT can be driven by constraints best described at the level of physical forms (e.g., the geospatial configuration and dynamics of an HMT in the shared work environment) as well as by constraints best described at the level of abstract functions, such as balances and trade-offs of energy, values, and priorities. Thus, an effective framework allows modeling such processes at multiple levels of abstraction to support identification and mapping of patterns in team behavior across multiple levels of abstraction. For example, through keeping track of relevant interdependencies in work across various levels of abstraction we can begin to identify common themes that lead to cascading failures in the system.

Level of aggregation refers to the detail in which the HMT is modeled within a framework. To enable modeling and evaluation at various stages in a development process (from early discovery with a notional idea of the system to having detailed descriptions of the DCWS), the framework should afford flexibility in the level of aggregation that is necessary to generate insightful results. The scalability of the system will in large part trend with upticks in fidelity, as increased ecological validity of the simulation will come hand in hand with increased sensor inputs, new agents, and other generators of information. Thus, the framework should not have a preconceived notion of the 'correct' level of detail.

DISCUSSION

The requirements described in this paper outline an ambitious plan to bridge technical gaps in the integration of humans, robots, and AI in joint activity. In doing so, we lay the theoretical groundwork for future quantitative analysis of HMT resilience and describe other pertinent themes relating to the evaluation of HMTs.

Future work will focus on developing a computational framework, using these requirements as a guide. These development efforts will build on existing simulation frameworks. For example, the modeling and simulation framework Work Models that Compute (WMC) has been successfully employed in investigating the dynamics of air traffic control and manned space operations (Feigh et al., 2001; IJtsma, 2019; Rosso & Saurin, 2018). By expanding existing simulation capabilities and developing new capabilities that capture principles of resilience engineering, we can begin to operationalize the architectural characteristics and governance protocols that lead to resilient behavior and develop actionable design guidance.

Examples of issues that can be explored with new computational modeling capabilities include:

1) Identifying and categorizing interdependencies present in multi-agent human robot teams
2) Generating hypotheses predicting simulated performance and performance in-situ as failures cascade in certain cases
3) Exploring connections between the selection of specific interdependencies in work as an explicit engineering tradeoff to bolster resilience

Modeling and simulation for research purposes can also be thought of as an act of reverse engineering, unpacking complex, multi-agent systems to understand what makes them resilient, and to then translate this understanding into guidelines for designing new capabilities.

For the final version of this paper we plan to further clarify the requirements and include a description of the current state of the art of agent-based modeling in resilience engineering, with a comparison of the identified requirements and this state-of-the-art.

CONCLUSIONS

The objective of this research effort is to develop new modeling and simulation capabilities for operationalizing key concepts of resilience in HMTs, with the aim of supporting developers and researchers in the analysis and creation of resilient HMTs. This paper outlined requirements for a framework aimed at operationalizing key concepts of resilience.

Resilience as described by Woods is thematically relevant to the evaluation of HMT performance, boasting a proactive approach to combat complexities associated with human-machine teaming. Specifically, sustained adaptability will likely be operationalized more fully in future work in this area, as evidence of resilient team behavior holds potential to exist as a meta-evaluative tool for measuring HMT stress-response and adaptive capacity.

REFERENCES

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Banks, A. P., & Millward, L. J. (2009). Distributed mental models: Mental models in distributed cognitive systems. *Journal of Mind and Behavior*, *30*(4), 249–266.

Cooke, N. J., Gorman, J. C., Duran, J. L., & Taylor, A. R. (2007). Team Cognition in Experienced Command-and-Control Teams. *Journal of Experimental Psychology: Applied*, *13*(3), 146–157. https://doi.org/10.1037/1076-898X.13.3.146

Cooke, N. J., Gorman, J. C., & Rowe, L. J. (2008). An ecological perspective on team cognition. *Team Effectiveness in Complex Organizations: Cross-Disciplinary Perspectives and Approaches*, *2173*, 157–182. https://doi.org/10.4324/9780203889312

Feigh, K. M., Denq, T. W., & Jacko, J. A. (2007). Contextual Control Modes During an Airline Rescheduling Task. *Journal of Cognitive Engineering and Decision Making*, *1*(2), 169–185. https://doi.org/10.1518/155534307X232839

Feigh, K. M., Pritchett, A. R., Jacko, J. a, & Denq, T. (2001). *Decision Making During an Airline Rescheduling Task: A Contextual Control Model Description*. 1–6.

Gervits, F., Thurston, D., Thielstrom, R., Fong, T., Pham, Q., & Scheutz, M. (2020). Toward genuine robot teammates: Improving human-robot team performance using robot shared mental models. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, *2020-May*, 429–437.

Hollnagel, E., & Woods, D. (2005). Joint Cognitive Systems. In *Joint Cognitive Systems*. https://doi.org/10.1201/9781420038194.ch6

Hollnagel, E., & Woods, D. D. (1999). Cognitive systems engineering: New wine in new bottles. *International Journal of Human Computer Studies*, *51*(2), 339–356. https://doi.org/10.1006/ijhc.1982.0313

IJtsma, M. (2019). *Computational Simulation of Adaptation of Work Strategies in Human-Robot Teams*. *August*.

IJtsma, M., Ye, S., Feigh, K., & Pritchett, A. (2019). Simulating Human-Robot Teamwork Dynamics for Evaluation of Work Strategies in Human-Robot Teams. *International Symposium on Aviation Psychology*, 103–108.

Johnson, M. (2014). *Coactive Design - Designing Support for Interdependence in Human-Robot Teamwork*. Delft University of Technology.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive Design: Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction*, *3*(1), 43. https://doi.org/10.5898/JHRI.3.1.Johnson

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common Ground and Coordination in Joint Activity. In *Organizational Simulation* (Issue September 2018). https://doi.org/10.1002/0471739448.ch6

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(6), 91–95. https://doi.org/10.1109/MIS.2004.74

Mingyue Ma, L., Fong, T., Micire, M. J., Kim, Y. K., & Feigh, K. (2018). *Human-Robot Teaming: Concepts and Components for Design*. 649–663. https://doi.org/10.1007/978-3-319-67361-5_42

Oishi, K., Gennari, J., Lewis, M., Hussain, A., Wang, J., & Yong, M. (2011). *A Hybrid Architecture for USAR Robot Development and Human User Testing*. 1–7.

Rosso, C. B., & Saurin, T. A. (2018). The joint use of resilience engineering and lean production for work system design: A study in healthcare. *Applied Ergonomics*, *71*(March), 45–56. https://doi.org/10.1016/j.apergo.2018.04.004

Schermerhorn, P., Kramer, J., Brick, T., Anderson, D., Dingler, A., & Scheutz, M. (2007). DIARC: A testbed for natural human-robot interaction. *AAAI Workshop - Technical Report*, *WS-06-15*, 45–52.

Vicente, K. (1999). *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. CRC Press.

Wang, J., Lewis, M., & Gennari, J. (2003). Usar: A Game Based Simulation for Teleoperation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(3), 493–497. https://doi.org/10.1177/154193120304700351

Wang, N., Pynadath, D. V., Unnikrishnan, K. V., Shankar, S., & Merchant, C. (2015). Intelligent agents for virtual simulation of human-robot interaction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9179*, 228–239. https://doi.org/10.1007/978-3-319-21067-4_24

Woods, D. D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering and System Safety*, *141*(September), 5–9. https://doi.org/10.1016/j.ress.2015.03.018

Woods, D. D. (2018). The theory of graceful extensibility: basic rules that govern adaptive systems. *Environment Systems and Decisions*, *38*(4), 433–457. https://doi.org/10.1007/s10669-018-9708-3

Woods, D. D. (2019). Essentials of resilience, revisited. *Handbook on Resilience of Socio-Technical Systems*, *January*, 52–65. https://doi.org/10.4337/9781786439376.00009

Woods, D. D., & Roth, E. M. (1994). Symbolic AI Computer Simulations as Tools for Investigating the Dynamics of Joint Cognitive Systems. *Expertise and Technology*, *January 1995*, 75–90.