14th Organizational Design and Management Conference
July 11-13, 2023, Bordeaux France

Future of work with artificial intelligence: designing distributed, joint, and self-organizing sociotechnical systems

Ashleigh Brady* and Neelam Naikar
Defence Science and Technology Group, Australia
*e-mail: ashleigh.brady1@defence.gov.au

Abstract
This paper aims to define critical properties of design models for human-AI systems that are informed by contemporary views of complex work performance. Recent disasters, such as the COVID-19 pandemic and wildfires in America, Europe, and Australia, serve as stark reminders that complex operations require collaboration among multiple dispersed teams working with a diverse collection of artifacts and networked technologies, including automation. Yet, prevailing models of human-automation interaction tend to be dyadic in nature, presupposing individual humans engaging with individual machines. Considering recent advances in artificial intelligence (AI) and the increasing interest among organizations to integrate these technologies in complex operations, we examine the concepts of distributed cognition, joint cognitive systems, and self-organization, which provide distinct perspectives of human-machine interactions in challenging settings. From these perspectives, we abstract critical ideas for the design of human-AI systems, and propose that design frameworks informed by contemporary views of complex work are needed.

Keywords
Human-machine systems, human-automation interaction, self-organization, distributed cognition, joint cognitive systems

Context
The increasing complexity of sociotechnical systems has been credited chiefly to large-scale computerization (Vicente, 2004). As computer-based systems have come into widespread use, shifting people's jobs and roles from physical to intellectual work, the ways in which these technologies may be deployed inadequately, and contribute to problems, failures, and accidents, has become increasingly apparent. A common lament is that technologies and systems are not designed with a human-centered approach, which implies that, by embracing such a viewpoint, numerous technological errors may be circumvented. However, many current models of human-automation interaction are dyadic in nature, focusing on interactions between a single human and a single machine. In addition, they tend to limit attention to individual, isolated tasks, devoid of the ecology and dynamic context in which work is situated, and also overlook requirements for communication, collaboration, and adaptive problem-solving. These models do not reflect contemporary views of complex work systems, and may be just as flawed as designs that prioritize technology over human needs. Moreover, as AI-enhanced machines become increasingly more capable, and are able to take on a widening array of complex tasks, these issues may become even more pronounced.

Objectives
Our objectives are to define critical properties of design models for human-AI systems that are informed by contemporary views of complex work systems.

Methods

Future of work with artificial intelligence: designing distributed, joint, and self-organizing cognitive systems

We examine the concepts of distributed cognition, joint cognitive systems, and self-organization, concentrating on the primary texts detailing these perspectives. Based on this analysis, we identify some of the defining themes of these concepts and consider the key implications for designing human-AI systems.

Main results

In his influential monogram *Cognition in the Wild*, Hutchins (1995) articulates the theory of distributed cognition and provides a compelling argument that viewing cognition as solely the property of individual minds is limiting. Instead, through careful, longitudinal observations of a US Navy ship navigating in enclosed waters, he demonstrates that computational processes involved in navigation are distributed across people, artifacts, and other aspects of the environment. These processes are spread in time and space, and actions taking place in one context can transform those occurring elsewhere. Moreover, the distributions may vary according to the circumstances. For example, Hutchins observed that when the ship's gyrocompass failed, the organizational structure of the navigation team changed considerably as a result. Hutchins's focus on cognition has the unfortunate consequence, among some audiences, of de-emphasizing the role of communication and coordination in complex work performance. However, a detailed reading of his work shows clearly that interactions among people and their artifacts are just as critical as those entities themselves. Thus, this perspective recognizes the importance of both individual minds and the interactions between people and their environments in cognitive work. By extending the boundaries of cognitive processes beyond the individual, broader interactions between internal and external processes can be considered. A key implication for the design of human-AI systems is the need to understand how computational work is distributed across multiple entities in time and space, and how coordination between those distributed elements can be supported.

The concept of a joint cognitive system put forward by Hollnagel and Woods (2005) also appears to promote a cognitive view of sociotechnical systems, albeit inadvertently (Hollnagel, 2022). In this view, a cognitive system is defined as one "that can modify its behavior on the basis of experience so as to achieve specific anti-entropic ends" (Hollnagel & Woods, 2005, p. 22), rather than one that is capable of processing information or computation. This perspective emphasizes the co-agency of humans and machines, as it is argued that neither humans nor machines act independently to achieve system goals. Accordingly, this view promotes a functional rather than structural account of human-machine systems. In other words, the physical separateness of humans and machines, though not denied, is ignored, and instead the human-machine system is viewed as a whole with the focus on its performance. Consequently, as well as viewing humans and technology as a functional unity working towards joint goals, the notion of co-agency places the focus on what the system does, rather than how it does it or what it is. In particular, emphasis is placed on understanding what it does when challenged with change and surprise to adapt and stay in control of the situation (Woods & Hollnagel, 2006). A primary implication of this perspective is that design should focus on considering what the human-machine entity does to meet the goals of the system, rather than on how interactions between physically separated humans and machines can be mediated through an interface.

The theory of self-organization has emerged from diverse fields, including cybernetics (Ashby, 1947, 1962; von Foerster, 1960) and thermodynamics (Nicolis & Prigogine, 1977), which has inevitably led to a range of terms, methods, and conceptualizations. Common to all notions of self-organizing systems is the absence of centralized control. Rather, control is characterized as being distributed over the whole system, with all parts contributing to the resulting organization (Heylighen, 2001). When considered in relation to sociotechnical systems, the concept of self-organization appears to provide a plausible framework for understanding how novel organizational forms appear to emerge spontaneously in response to changes in the system's circumstances (Naikar, 2018; Naikar & Elix, 2021). Specifically, this perspective posits that, outside of routine or recurring conditions, there may be situations where a system's formal organizational structure constrains its response in ways that are unsuitable or unproductive. As a result, individual, interacting actors may spontaneously adapt their behaviors to meet the demands of the situation. These spontaneous behaviors and interactions may result in changes in the work organization that are better suited to the local circumstances, so that the emergent organization stabilizes, and continues to constrain and enable actors' behaviors under those conditions. However, in response to ongoing changes in the situation, actors' organizational structures and behaviors may evolve continuously in a relatively seamless—if not perfect—self-organizing process. This persistent interaction of individual behavior with system organization provides the flexibility and coordination needed in the sociotechnical system to function effectively in dynamic environments, even when faced with high levels of uncertainty about the current situation as well as unpredictability in the future direction of events (Naikar, 2020). The primary implication of this perspective for the design of human-AI systems is to preserve or maximize the system's possibilities for action, or requisite variety of the system (Ashby, 1956), so that it can continuously adapt its behavior and organization to stay in control of unfolding events.

Discussion/perspectives

The contemporary perspectives of distributed cognition, joint cognitive systems, and self-organization, though distinct, view humans and machines through the common lens of an integrated system. Also well known is that these perspectives have the shared motivation of understanding how work is carried out in complex operational settings, recognizing that it does not involve an isolated human and machine performing a single task, but rather multiple humans and machines engaged in a rich variety of interconnecting activities. However, it appears to be less widely appreciated that all three perspectives also acknowledge the important role that adaptation plays in effective system functioning in complex environments—in particular, this aspect of Hutchins's (1995) work has received much less attention than his observation that cognition is best conceived as distributed rather than isolated within an individual's mind (Naikar, 2020).

Nevertheless, a number of potential discrepancies between the three perspectives are also evident. One discrepancy relates to the boundaries applied to human-machine systems. The theory of distributed cognition extends the traditional boundaries of cognitive work to incorporate social, physical, and temporal spaces, thereby accounting for interactions between people and their environments. Similarly, in the self-organization perspective, the ecology is considered a critical part of the work system, as it both limits and affords possibilities for action from which coordinated work patterns may be constructed. However, Norros and Salo (2009) have argued that it is unclear whether the environment in which humans and technologies are situated is included as part of the joint cognitive system. For example, they point out that, in Woods and Hollnagel's (2006) coverage on the patterns of affordance, the relationships captured appear to be restricted to those between humans and technologies; in other words, the affordances of the environment are overlooked.

A second area of incompatibility relates to the extent to which humans and machines can be seen as distinct entities. The joint cognitive systems perspective, while not denying the physical separateness of humans and machines, argues that it is more important to see humans and machines as a functional unity. In contrast, by recognizing the distribution of work processes across social, physical, and temporal spaces, the theory of distributed cognition emphasizes distinct entities such as people, artifacts, and other structures in the environment. The self-organization perspective accommodates both points of view. Specifically, it recognizes that system coherence is achieved, or emerges, from the spontaneous actions of individual, interacting, actors in the environment they cohabit, which both affords and limits possibilities for action.

Finally, a third area of discrepancy relates to whether cognitive or computational processes of actors is considered. Clearly, the theory of distributed cognition emphasizes the cognitive and computational processes of actors. In contrast, the joint cognitive systems perspective places the focus on understanding 'what the system does', not 'how it does it'. The self-organization perspective recognizes that the cognitive and computational capacities of humans and machines, respectively, allow for their spontaneous actions and interactions, which form the basis for emergence in the work organization.

To the extent that these perspectives are empirically substantiated or logically argued, they may each be seen as offering a useful vantage point for design. Moreover, when considered collectively, they lead to a number of critical considerations for the design of human-AI systems. In particular, whereas the distributed cognition perspective emphasizes the distribution of work across multiple entities, and the joint cognitive systems concept emphasizes the functional unity of these entities, the self-organization perspective explains how system coherence can emerge from the spontaneous actions of multiple, distributed elements, providing the system with the adaptive capacity for staying in control of the situation. We therefore propose that work in human-AI systems must be organized in a way that enables coordinated performance to emerge from the boundaries imposed by the environments in which humans and machines are distributed and the competencies of these actors. Designs that fail to consider the environment in which work is performed, and the capacities of actors for the work, bounds the system in a way that risks overlooking significant interactions between these elements and artificially limits the system's space of possibilities for action. Consequently, the system's adaptive capacity may be compromised.

Models for design are needed that reflect contemporary advances in our understanding of complex work systems. Conventional human-automation interaction models tend to focus on individual humans and individual machines, reducing the problem to simple, dyadic components and relationships. Systems, however, are more than the sum of their parts in that interactions within systems may give rise to properties or behaviors that cannot develop in the component parts (von Bertalanffy, 1968).

While the concepts of distributed cognition, joint cognitive systems, and self-organization focus on providing explanatory accounts of complex work performance, the joint cognitive systems approach incorporates a functional perspective that informs design more directly. However, translating insights from observed patterns of work into functional designs has proved challenging for communities relying on such perspectives and remains largely undemonstrated (Righi, Saurin, & Wachs, 2015). Moreover, taking into account the insights of all three perspectives, we propose that frameworks for design are needed that consider not only what the human-machine system does, but also the environment and functional capacities of human and machine actors, in a way that can account for emergence of new work patterns and system capability through self-organization.

We suggest that cognitive work analysis is a formative framework (Rasmussen, Pejtersen, & Goodstein, 1994; Vicente, 1999) with this potential. For example, its first three dimensions (work domain analysis, activity analysis, and strategies analysis), model the work demands of the system independently of whether the actors are humans or artificial agents, accommodating the view of people and technology as parts of a whole that are capable of jointly performing work. In addition, the first two dimensions account for the boundaries of the physical and social ecology (Naikar, 2013) and the temporal and spatial properties of work situations (Naikar, Moylan, & Pearce, 2006), respectively. This approach is consistent with the theory of distributed cognition, which suggests that the boundaries of cognitive processes must extend beyond the individual to their social, physical, and temporal spaces to account for interactions between people and their environments. Third, the diagram of work organization possibilities (Naikar & Elix, 2016), which is used in the social organization and cooperation dimension of cognitive work analysis, accommodates the emergence of novel work patterns from the spontaneous actions of individual, interacting actors, whether these are humans or machines. It therefore offers an approach to the design of human-AI systems that does not artificially limit a system's capacity for adaptation, and hence its capacity to operate successfully in dynamic, uncertain, and unpredictable circumstances (Naikar, 2018).

In conclusion, in this paper we have examined the defining themes of distributed cognition, joint cognitive systems, and self-organization to abstract critical considerations for the design of human-AI systems. Through our examination of the central ideas of these perspectives, we have learned that, although they share a common view of humans and machines as essential components of an integrated system, they also exhibit marked discrepancies. However, to the extent that these perspectives are grounded in empirical observations and logical arguments, they can each offer useful frames of reference for design. Therefore, considering these perspectives collectively, we have arrived at the proposition that the design of systems with humans and machines must facilitate the emergence of coordinated work patterns within the boundaries of their environments and respective competencies. We have also put forward the notion that cognitive work analysis provides a principled framework for integrating important insights from the three perspectives in a way that is useful for design. Nevertheless, considerable further work is required to investigate the full extent to which the key ideas of distributed cognition, joint cognitive systems, and self-organization, and the compatibilities and discrepancies between these perspectives, can be handled effectively by cognitive work analysis.

References

Ashby, W. R. (1947). Principles of the self-organizing dynamic system. *The Journal of General Psychology*, *37*, 125–128. https://doi.org/10.1080/00221309.1947.9918144

Ashby, W. R. (1956). An introduction to cybernetics. John Wiley & Sons. https://doi.org/10.5962/bhl.title.5851

Ashby, W. R. (1962). Principles of the self-organizing system. In H. von Foerster & G. W. Zopf Jr. (Eds.), *Principles of self-organization* (pp. 255-278). Pergamon Press.

Heylighen, F. (2001). The science of self-organization and adaptivity. *The Encyclopedia of Life Support Systems*, *5*(3), 253-280.

Hollnagel, E. (2022, n.d.). *Cognitive systems engineering: RIP*. Erik Hollnagel. https://erikhollnagel.com/ideas/cognitive-systems-engineering.html

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press. https://doi.org/10.1201/9781420038194

Hutchins, E. (1995). *Cognition in the wild.* MIT Press.

Naikar, N. (2013). *Work domain analysis: Concepts, guidelines, and cases.* CRC Press. https://doi.org/10.1201/b14774

Naikar, N. (2018). Human–automation interaction in self-organizing sociotechnical systems. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 62–66. https://doi.org/10.1177/1555343417731223

Naikar, N. (2020). Distributed cognition in self-organizing teams. In M. McNeese, E. Salas, & M. Endsley (Eds.), *Contemporary research: Models, methodologies, and measures in distributed team cognition*. CRC Press. https://doi.org/10.1201/9780429459733

Naikar, N., & Elix, B. (2016). Integrated system design: Promoting the capacity of sociotechnical systems for adaptation through extensions of cognitive work analysis. *Frontiers in Psychology*, *7*(962), 44-64. https://doi.org/10.3389/fpsyg.2016.00962

Naikar, N., & Elix, B. (2021). Designing for self-organisation in sociotechnical systems: Resilience engineering, cognitive work analysis, and the diagram of work organisation possibilities. *Cognition, Technology & Work, 23*, 23-37. https://doi.org/10.1007/s10111-019-00595-y

Naikar, N., Moylan, A., & Pearce, B. (2006). Analysing activity in complex systems with cognitive work analysis: Concepts, guidelines and case study for control task analysis. *Theoretical Issues in Ergonomics Science*, *7*(4), 371-394. https://doi.org/10.1080/14639220500098821

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations.* Wiley. https://doi.org/10.1002/bbpc.197800155

Norros, L., & Salo, L. (2009). Design of joint systems: A theoretical challenge for cognitive systems engineering. *Cognition, Technology & Work*, *11*, 43-56. https://doi.org/10.1007/s10111-008-0122-3

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. John Wiley & Sons. https://doi.org/10.1002/(SICI)1099-1360(199603)5:1<75::AID-MCDA92>3.0.CO;2-4

Righi, A. W., Saurin, T. A., & Wachs, P. (2015). A systematic literature review of resilience engineering: Research areas and a research agenda proposal. *Reliability Engineering and System Safety*, *141*, 142–152. https://doi.org/10.1016/j.ress.2015.03.007

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* Lawrence Erlbaum Associates. https://doi.org/10.1201/b12457

Vicente, K. J. (2004). *The human factor: Revolutionizing the way people live with technology*. Routledge. https://doi.org/10.4324/9780203944479

von Bertalanffy, L. (1968). *General system theory: Foundations, development, applications.* George Braziller.

von Foerster, H. (1960). On self-organizing systems and their environments. In M. C. Yovitts & S. Cameron (Eds.), *Self-organizing systems* (pp. 31-50). Pergamon.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering.* CRC Press. https://doi.org/10.1201/9781420005684

