Towards Theoretical Foundations for Human-Autonomy Teams

Marc Steinberg, Office of Naval Research

I. Introduction

Bringing systems-level theoretical foundations to the design and development of a Human Autonomy Team (HAT) has many challenges compared to a more traditional division of human and machine roles, responsibilities, and functions (Klein, 2004, Groom, 2007, Shah, 2010, Cooke, 2013, Gao, 2016, Endsley, 2016, McNeese, 2018, Johnson, 2019). HAT may involve new types of organizational structures in which multiple humans dynamically interact with multiple autonomous systems outside of fixed control hierarchies and with dynamically changing roles. Interaction between teammates may involve multi-modal tiered strategies with both verbal and non-verbal and explicit and implicit communications. Effective joint communication, attention, and action may depend on the ability to recognize individual capabilities, activities and status, and infer other team members' intent, beliefs, knowledge, and plans. Team activities may not be limited to just real-time task performance, but include also the ability to jointly train, rehearse, plan, and make a priori agreements prior to performing work together, and to assess performance and improve together afterwards. While this might appear a daunting list of capabilities to achieve in machines, it is not necessary that HAT operate on exactly the same principles as high functioning human teams that exhibit these characteristics. A much broader spectrum of group types is possible that could be considered teams, and would be more plausible to engineer in the near future. Further, the true value of HAT may lie in exploiting the heterogeneity between humans and machines to create entirely new types of organizations rather than trying to mimic fully human ones or force humans into the rigid frameworks of multi-agent machine systems. In this spirit, a human-autonomy team will be categorized in this article as requiring only the following properties:

1) Teams are set up to achieve a common goal or goals that are believed achievable in a bounded period of time. It is not required that every member has the same depth of understanding of the goal. This would be very challenging for machines on many complex, real-world problems, and is not the case for teams of humans and working animals or teams of human adults and children that may provide useful inspiration for the degree of heterogeneity to be found in HAT. As well, this is consistent with broader findings in the human team literature, particularly for teams that are heterogeneous or that have larger numbers of members (Cooke, 2013).

2) Teams exploit role specialization and have bi-directional interdependencies between teammates. Teaming interdependencies are not predominantly one-way, such as in human supervisory control of autonomy. Methods that focus primarily on decomposing and allocating loosely coupled tasks between humans and machines to ensure task completion with non-interference between agents would also not be sufficient on their own to be considered teaming.

3) Individual identities of teammates matter. This allows for unique relationships or associations to be formed between particular pairings or subsets of teammates. This differs from multi-agent forms of organization in which individual agents can be anonymous such as biologically-inspired collective behaviors (Steinberg, 2011), and call center or service oriented models with a pool of autonomous systems (Lewis, 2011).

To accomplish even just this for HAT goes beyond current systems theories or the methods of any particular discipline. Thus, it makes sense to consider foundations from as broad a perspective as possible. This article will consider a sampling of perspectives across scientific fields such biology, neuroscience, economics/game theory, and psychology, methods for developing and accessing complex socio-technical systems from human factors and organizational psychology, and methods for engineering teams from computer science, robotics, and engineering.

II. Organizational Structure and Role/Function Allocation

Groups of humans and working social animals have particular relevance for Human-Autonomy Teaming (HAT) because they encompass some of the same degree of extreme heterogeneity of physical, sensing, communication, and cognitive abilities. Additionally, recent animal cognition research has focused on the extent to which different animal species may excel at solving specific niches of cognitive problems under particular ecological constraints while being rather poor at others (Rogers & Kaplan, 2004). For example, there is an increasing body of evidence on the impressive social cognitive abilities dogs can use to solve problems jointly with humans, while simultaneously finding dogs can be much less capable at individually solving other classes of cognitive problems (Hare, 2013). This has similarities to the state of the art of today's autonomous systems, and may provide both an inspiration for teaming architectures and an effective metaphor for human interaction with autonomous teammates. There are a number of systemic frameworks from the animal literature that can be considered for HAT including different subordinate strategies (Sun, 2010), mutualism (Madden, 2010), or reciprocal altruism and association strengths between individuals (Haque, 2009). A dominance framework, for example, can provide a principled framework to allow for more freedom of action by human teammates, while limiting machine teammates to act within constraints imposed by human plans and actions. From an engineering perspective, this can be considered a satisficing type of solution. The constraints imposed by dominance relationships between team members both ensure some degree of non-interference, and also can substantially simplify computationally intractable group coordination problems so they can be solved even for complex groups at scale. For example, adopting a dominance like structure has enabled solution of large scale group problems with Decentralized Partially Observable Markov Decision Processes (Sun, 2010), and several approaches for motion planning with large numbers of systems in complex environments have achieved scalability with related types of prioritization and constraint (Herbert, 2017). There also have been a number of successes in showing how particular architectures relate to the degree of optimality, robustness, resilience, or the best or worst possible cases (Ramaswamy, 2019).

A different set of methods can be drawn from human factors. For example, in an assessment of the literature, Roth et al (2019) identified a four-stage process for role allocation in HAT to analyze operational and task demands, consider ways of distributing work across human and machine team members, examine interdependencies in both nominal and off-nominal conditions, and explore the trade space of options with different potential tools. One of the particular tools that has had success at such novel problem domains is Cognitive Work Analysis (CWA) (Vicente, 1999). CWA has been successfully applied to two related classes of problems of human supervisory control of autonomous teams in which the human is not a teammate (Linegang, 2006, Hoffman, 2008), and to the development of assistive technologies for human teams in fields like healthcare and aviation in which the automation is not a teammate (Ashoori, 2013). A strength of CWA for novel systems is that it is based on an ecological theory in which human/machine activity and interaction can be considered from the perspective of constraints on what is and is not possible in the work environment rather than starting with stronger assumptions on how the work will be done. Thus, it has been particularly effective on problems that are dominated by persistent fundamental constraints of physics or information flow. CWA could be effective for HAT problems with similar characteristics. However, there is only a limited body of work on extending the abstractions involved to team problems even in the fully human case. Further, there are many challenges in applying this kind of method, and some prior work has found that results do not sufficiently encompass what is enabled by the new technological options. Another method of considering interdependencies that was developed more specifically for HAT problems is Co-Active Design (Johnson, 2014).

Similar to human factors, human robotic interaction can provide a rich set of theories for human-autonomy coordination and adaptation that take into account the realities of implementing methods on real autonomous systems. Discussions of human-robot and human-autonomy teams sometimes proceed from the assumption that the machine team members will be something like a peer. However, realistically, many machine teammates in the next few decades will probably still require some degree of human supervision and support. This may be due to both technological limitations and interrelated issues such as laws, regulations, organizational policies, ethical concerns, societal norms, and professional standards of due diligence. Thus, some systems frameworks developed for non-team interactions with robotic and autonomous systems will still have validity. For example, an important framework for considering human-robotic interaction is span of control (Crandall, 2005). Historically, the focus of this has been on matching the human capacity to the tempo and quantity of "servicing" that the machines require. In one approach, this is based around a neglect time representing the amount of time a robot can operate safely and effectively or be trusted to do so without human intervention. In moving from an operator to a teammate, this framework may need to consider something more like a human sports team or a medical team. A given size and complexity of team might require a certain number of on-field leaders, and off-field coaches, trainers, and health and safety monitors. A converse of neglect time is neglect benevolence (Walker, 2012). Neglect benevolence recognizes that there are circumstances in which a lower bandwidth of interaction between some group members would be beneficial for team performance, including in the human to machine direction.

III. Working Together on Tasks

A sophisticated capability that is likely foundational for achieving higher performing teams is Theory of Mind (ToM). ToM is the ability to infer that others have different knowledge, beliefs, desires, and intentions than one's self. Neuroscience research on simultaneous imaging of multiple brains has found connections as well with how synchronization of behavior, language, and gesture is achieved in some types of group interactions (Dumas, 2010). For autonomous teammates, ToM could provide principled connections between perception, perceptual attention and active sensing, intent and activity recognition, knowledge and world representation, prediction, and decision-making in groups. Theory of Mind has been shown to exist in some form in increasingly younger human children (Doherty, 2009), and there are debates on the extent to which at least rudimentary ToM occurs in non-human primates, other mammals, and even several bird species (Rogers & Kaplan, 2004). An important divergence between psychological models and robotics has been that the three main scientific theories centered on an ability to project one's own experiences and way of thinking onto others. Autonomous systems, lacking remotely comparable brains and experiences to their human teammates, instead may need to have their ToM more heavily grounded in processes of observation and learning. Thus, robotic versions of ToM have tended to focus on narrower abilities of perspective taking, belief management with limited numbers of entities and objects, and bounded rationality that can be tailored for a particular experiment, but are difficult to scale to more realistic, open world problems (Scassellati, 2002, Brezeal, 2009, Hiatt, 2011, Weerd, 2013). However, scientific research has also begun to emphasize the role of observation in the natural development of such abilities, and this may be an excellent opportunity to reconsider ToM as a foundational theory for HAT.

Another important set of theories are those for joint action/activity (Clark, 1996, Bradshaw, 2009) and common ground (Stubbs, 2007). Joint activity involves the ability to coordinate tasks with interdependencies, and can depend on common ground as a kind of floor of the minimum knowledge, beliefs, and assumptions that are required to be shared and maintained between agents. This can range from direct communications between teammates to generally shared world knowledge or widely accessible broadcasts. For teams, this requires both regular updating and maintenance and the ability to recognize when it has broken down and needs to be repaired. Note that while common ground might seem to be a particularly human ability that is not the case. For example, dogs are capable of both spontaneously picking up on human cues and on signaling themselves in ways that can support joint problem solving with humans. Common ground via non-direct communication has connections to both the biological literature on stigmergy, in which coordination is done via changes in the environment (Steinberg, 2011) and Dynamic Epistemic Logic, which can be used to reason about changes of belief and knowledge that occur due to trustworthy announcements to a group (Lutz, 2006). There are also relationships of these concepts to game theory research on how agents adapt to each other under some degree of bounded information (Fudenberg, 1998). In the study of human teams, joint action models often revolve around some notion of an agreement between agents that need to be maintained along with common ground. However, within the biological and economics/game theory literature there are debates on the extent to which seemingly strongly coordinated activities can instead arise as the result of more decentralized decision-making (Madden, 2010, Young, 2014). Common ground and joint activity has sometimes been interpreted in robotics as the kind of information and plan representations found already in robotics "world models" or on operator displays and there are a number of methods that have been considered for different aspects of this such as information theory, Partially Observable Markov Decision Processes (POMDP), bounded rationality, and Hierarchical Task Networks (HTN) (Roth, 2005, Unhelkar, 2016). However, these are much broader concepts, and an important part of common ground in HAT will be bridging the divide between human and machine representations of goals, tasks, and understanding of the world and each other. Relating to joint activity in Human Robotic Interaction is also the idea of shared control (Mulder, 2015). Some shared control research has focused very application specific designs, such as human assistance tele-operations problems. However, other research can generalize this in a more abstract ways that may be a good model for some of the richer interactions that may occur in HAT. For example, one approach is to provide a systematic way for a machine teammate to estimate the quality of interactions it is having with a human relative to goal achievement and then to be able to adjust the level of interaction and allocation of tasks to one appropriate to the circumstances (Javaremi, 2019). A second example of shared control uses a formal Linear Temporal Logic approach as a way to reason about shared policies (Fu, 2016). Other examples utilize an optimal control paradigm that recognizes differences in machine and human understanding of the problem for physical tasks, and addresses the issue of how humans may adapt to autonomy over time (Nikolaidis, 2017) or takes into account a distribution of potential human goals if the actual human goal is unknown (Javdani, 2018).

Another group of methods from human factors and organizational psychology were developed specifically for human teams. A traditional approach has been models of shared cognition such as team mental models and shared mental models (Lim, 2006, Mohammed, 2010). A mental model, in this case, is a representation that allows the behavior of a system to be described, explained, and predicted. This group of theories provides a way to aggregate that as an information structure across the group. Shared mental models have also been a popular idea in human robotic interaction and related forms of AI and autonomy, but the mechanizations of these often is narrowly tailored to particular problems compared to the versatility that is implied in the human case and often focuses on awareness rather than comprehension and prediction. There are challenges as well to deal with the heterogeneity of HAT. For example, in an ad hoc team, broad knowledge held by a machine may be less likely to be available, recognized as relevant, shared, or acted upon in a timely way than in a human team member. The differing nature of the human machine/interactions may play a more significant role than the team information structure. An alternative approach is Integrated Team Cognition (Cooke, 2013). This approach arguably has significant compatibility with engineering and computer science methods in that it takes a bottom up, layered, dynamical systems approach based on observable interactions. This approach has been applied to small human-autonomy teams with sophisticated synthetic team members based on a full cognitive architecture. Research also has included off-nominal performance, failures, and compromising of the autonomous teammate (McNeese, 2018, Gorman, 2019). Other important classes of methods from human factors include theory-based approaches to situation awareness (Endsley, 2000), transparency (Chen, 2018), and trust (Lee, 2004, Hancock, 2011) that have been effective on related classes of problems and applied to some cases of HAT. A related concept from human robotic interaction is that of legibility and predictability (Dragan, 2013). Legibility represents how well an observer could rapidly infer the system's goals from observed behavior while predictability relates to the extent that observed behavior is what would be expected given a known goal.

A final related area concerns models of emotion, affect, and motivation, how these may vary among individuals and relate to interaction and communication among team members. There has been considerable growth in the development of cognitive and neuroscience models of the role of affect and motive in cognition and even some principled systems theories in robotics based on either psychological or neural models (Moshkina, 2011). However, much research in this area has focused on problems such as virtual training environments, tutoring systems, games, toys, artificial pets, and companions. Some cognitively plausible models of affect and motivation have been applied to assist robots in their ability to communicate to humans in social domains. However, this research has often focused on the ability of the autonomous system to provide a more pleasant experience for the user, improved communications, and usability rather taking a more functional perspective towards being an effective teammate that performs tasks with humans towards a common goal.

III. Teaming over Longer Durations

Much research to date on Human-Autonomy Teams has focused narrowly on relatively small teams performing short time duration tasks. Nonetheless, creating effective autonomous teammates must also consider aspects like the joint training of both machine and human members of the team, the ability for the team to jointly do pre-task planning, agreements, and rehearsal, and post-task assessment, maintenance, and improvement. Human factors can provide both general frameworks to support the design, development, and analysis of complex socio-technical systems including some of the methods described above. These have the advantage of encompassing a broad range of Human Systems Integration concerns, but can require a great deal of care and creativity to extend to a fundamentally novel concept like a Human-Autonomy Team (HAT). For example, joint training and rehearsal of both humans and machines has not had much study. However, there are theories that exist with regards to human training with autonomy (Zhou, 2019), autonomy as tutors for humans, and frameworks for interactive machine learning in which humans assist machines in learning. The latter has had some work that has considered human factors and human centered design aspects that go beyond treating the human mainly as servicing the automation (Krening, 2018). Methods like shielding potentially provide a verifiable way to ensure machine learning stays within safe bounds around other teammates. Joint planning also has had some work that has considered both human factors and human models within planning and risk management. An approach towards pre-task agreements that also has potential value for verification and decomposition of HAT is contract-based approaches (Nuzzo, 2015, Benveniste, 2018). At a systems level, these can be used to guarantee global properties as long as each individual element abides by a set of guarantees that are rooted in local assumptions. In the event that an assumption is violated, there is research on monitoring and adapting contracts to be able to restore some guarantees in real-time. This might seem like a very difficult method to bridge across people and machines, but some similar kinds of approaches have been successful with people. Finally, another significant area is self-assessment and prediction of proficiency and the ability to communicate this effectively to human teammates in terms of achievable performance over a range of operating parameters prior to starting a task, in real-time while performing a task, and then afterwards using knowledge of the completed task (Steinfeld, 2019).

Conclusions and Future Directions

Creating foundational systems theories for Human Autonomy Teams (HAT) raises new issues that are substantially different from those that have previously been encountered in related areas such in the study of fully human teams or of human management and supervision of fully machine teams. This article has discussed a diverse set of generalizable theories from different disciplinary communities that could be part of that basis for HAT. The different communities involved each bring different strengths and weaknesses and often ask different kinds of questions and make different kinds of assumptions. Further, there has been some reduction in the amount of focus on foundational theories in recent years within some of these disciplines due to the popularity of data-driven machine learning methods. Central to the needs of many of these methods are the extension of formal, cognitive, and biologically plausible models of humans from ones that can predict human performance only within narrowly defined problems to ones that can accommodate the less structured and richer decision, action, and interaction environments of highly heterogeneous and adaptive teams. For example, there has been considerable progress in learning useful models of human driver behavior as part of research on autonomous cars, but this often involves a fairly narrow set of short time duration modeling problems that wouldn't extend to team problems involving longer duration complex sequential decision-making problems with significant dependencies, uncertainties, and criticalities across the team. If advances in the underlying human models and theories can be made, than many different methods from engineering and computer science that assume such predictive models may become viable in support of HAT. If the advances in human models are not forthcoming, than progress towards systematic foundations is likely to be slower and perhaps require substantially new thinking.

References

Ashoori, M., & Burns, C. (2013). Team cognitive work analysis: Structure and control tasks. *Journal of Cognitive Engineering and Decision Making*, 7(2), 123-140.

Bradshaw, J. M., Feltovich, P., Johnson, M., Breedy, M., Bunch, L., Eskridge, & van Diggelen, J. (2009, July). From tools to teammates: Joint activity in human-agent-robot teams. In *International conference on human centered design* (pp. 935-944). Springer, Berlin, Heidelberg.

Breazeal, C., Gray, J., & Berlin, M. (2009). An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research*, *28*(5), 656-680.

Benveniste, A., Caillaud, B., Nickovic, D., Passerone, R., Raclet, J. B., Reinkemeier, P., ... & Larsen, K. G. (2018). Contracts for system design. *Foundations and Trends in Electronic Design Automation*, 12(2-3), 124-400.

Chen, J. Y., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science*, *19*(3), 259-282.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, 37(2), 255-285.

Crandall, J. W., Goodrich, M. A., Olsen, D. R., & Nielsen, C. W. (2005). Validating human-robot interaction schemes in multitasking environments. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 35(4), 438-449.

Doherty, M. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Psychology Press.

Dragan, A. D., Lee, K. C., & Srinivasa, S. S. (2013, March). Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 301-308). IEEE.

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PloS one*, 5(8).

Endsley, M. R., & Garland, D. J. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement*, *1*, 24.

Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. *Human factors*, *59*(1), 5-27.

Fu, J., & Topcu, U. (2015). Synthesis of shared autonomy policies with temporal logic specifications. *IEEE Transactions on Automation Science and Engineering*, *13*(1), 7-17.

Fudenberg, D., Drew, F., Levine, D. K., & Levine, D. K. (1998). *The theory of learning in games* (Vol. 2). MIT press.
Gao, F., Cummings, M. L., & Solovey, E. (2016). Designing for robust and effective teamwork in human-agent teams. In *Robust intelligence and trust in autonomous systems* (pp. 167-190). Springer, Boston.
Gorman, J. C., Demir, M., Cooke, N. J., & Grimm, D. A. (2019). Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: a layered dynamics approach. *Ergonomics*, *62*(5), 629-643.
Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, *8*(3), 483-500.
Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5), 517-527.
Haque, M. A., & Egerstedt, M. (2009, June). Coalition formation in multi-agent systems based on bottlenose dolphin alliances. In *2009 American Control Conference* (pp. 3280-3285). IEEE.
Hare, B., & Woods, V. (2013). *The genius of dogs: How dogs are smarter than you think*. Penguin.
Herbert, S. L., Chen, M., Han, S., Bansal, S., Fisac, J. F., & Tomlin, C. J. (2017, December). FaSTrack: A modular framework for fast and guaranteed safe motion planning. In 2017 IEEE 56th Annual Conference on Decision and Control (CDC) (pp. 1517-1522). IEEE.
Hiatt, L. M., Harrison, A. M., & Trafton, J. G. (2011, June). Accommodating human variability in human-robot teams through theory of mind. In *Twenty-Second International Joint Conference on Artificial Intelligence*.
Hoffman, J. D., Lee, J. D., & Seppelt, B. D. (2008, September). Identifying display concepts to support distributed collaboration of unmanned vehicle teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 52, No. 5, pp. 488-492). Sage CA: Los Angeles, CA: SAGE Publications.
Javaremi, M. N., Young, M., & Argall, B. D. (2019, June). Interface Operation and Implications for Shared-Control Assistive Robots. In 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR) (pp. 232-239).
Javdani, S., Admoni, H., Pellegrinelli, S., Srinivasa, S. S., & Bagnell, J. A. (2018). Shared autonomy via hindsight optimization for teleoperation and teaming. *The International Journal of Robotics Research*, 37(7), 717-742.
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), 43-69.
Johnson, M., & Vera, A. (2019). No AI is an island: The case for teaming intelligence. *AI Magazine*, *40*(1), 16-28.
Klien, 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.
Krening, S., & Feigh, K. M. (2018). Interaction algorithm effect on human experience with reinforcement learning. ACM Transactions on Human-Robot Interaction (THRI), 7(2), 1-22.
Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80.
Lewis, M., Wang, H., Chien, S. Y., Velagapudi, P., Scerri, P., & Sycara, K. (2011). Process and performance in human-robot teams. *Journal of Cognitive Engineering and Decision Making*, 5(2), 186-208.
Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *27*(4), 403-418.
Linegang, M. P., Stoner, H. A., Patterson, M. J., Seppelt, B. D., Hoffman, J. D., Crittendon, Z. B., & Lee, J. D. (2006, October). Human-automation collaboration in dynamic mission planning: A challenge requiring an ecological approach. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 23, pp. 2482-2486). Sage CA: Los Angeles, CA: SAGE Publications.
Lutz, C. (2006, May). Complexity and succinctness of public announcement logic. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems* (pp. 137-143).
Madden, J. D., Arkin, R. C., & MacNulty, D. R. (2010, December). Multi-robot system based on model of wolf hunting behavior to emulate wolf and elk interactions. In *2010 IEEE International Conference on Robotics and Biomimetics* (pp. 1043-1050). IEEE.

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. Human factors, 60(2), 262-273.
Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of management*, *36*(4), 876-910.
Moshkina, L., Park, S., Arkin, R. C., Lee, J. K., & Jung, H. (2011). TAME: Time-varying affective response for humanoid robots. *International Journal of Social Robotics*, *3*(3), 207-221.
Mulder, M., Abbink, D. A., & Carlson, T. (2015). Introduction to the special issue on shared control: applications. *Journal of Human-Robot Interaction*, *4*(3), 1-3.
Nikolaidis, S., Zhu, Y. X., Hsu, D., & Srinivasa, S. (2017, March). Human-robot mutual adaptation in shared autonomy. In 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI (pp. 294-302). IEEE.
Nuzzo, P., Sangiovanni-Vincentelli, A. L., Bresolin, D., Geretti, L., & Villa, T. (2015). A platform-based design methodology with contracts and related tools for the design of cyber-physical systems. Proceedings of the IEEE, 103(11), 2104-2132.
Ramaswamy, V., Paccagnan, D., & Marden, J. R. (2019, June). Multiagent maximum coverage problems: The trade-off between anarchy and stability. In 2019 18th European Control Conference (ECC) (pp. 1043-1048). IEEE.
Rogers, L. J., & Kaplan, G. (Eds.). (2012). *Comparative vertebrate cognition: are primates superior to non-primates?*. Springer Science & Business Media.
Roth, M., Simmons, R., & Veloso, M. (2005, July). Reasoning about joint beliefs for execution-time communication decisions. In Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems (pp. 786-793).
Roth, E. M., Sushereba, C., Militello, L. G., Diiulio, J., & Ernst, K. (2019). Function Allocation Considerations in the Era of Human Autonomy Teaming. Journal of Cognitive Engineering and Decision Making, 13(4), 199-220.
Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, *12*(1), 13-24.
Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human factors*, *52*(2), 234-245.
Steinberg, M. (2011, May). Biologically-inspired approaches for self-organization, adaptation, and collaboration of heterogeneous autonomous systems. In *Defense Transformation and Net-Centric Systems 2011* (Vol. 8062, p. 80620H). International Society for Optics and Photonics.
Steinfeld, A. et al, SUCCESS MURI: *Self-Assessment and Understanding of Conditions to Ensure System Success*, retrieved Dec. 3, 2019, site: https://successmuri.org/
Sun, X., Ray, L. E., Kralik, J. D., & Shi, D. (2010, October). Socially augmented hierarchical reinforcement learning for reducing complexity in cooperative multi-agent systems. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3244-3250). IEEE.
Unhelkar, V. V., & Shah, J. A. (2016, March). Contact: Deciding to communicate during time-critical collaborative tasks in unknown, deterministic domains. In *Thirtieth AAAI Conference on Artificial Intelligence*.
Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. CRC Press.
Walker, P., Nunnally, S., Lewis, M., Kolling, A., Chakraborty, N., & Sycara, K. (2012, October). Neglect benevolence in human control of swarms in the presence of latency. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3009-3014). IEEE.
De Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. Artificial Intelligence, 199, 67-92.
Young, P., & Zamir, S. (Eds.). (2014). *Handbook of game theory*. Elsevier.
Zhou, J., Zhu, H., Kim, M., & Cummings, M. L. (2019). The Impact of Different Levels of Autonomy and Training on Operators' Drone Control Strategies. ACM Transactions on Human-Robot Interaction (THRI), 8(4), 1-15.