Human Factors and Simulation, Vol. 83, 2023, 25–35
https://doi.org/10.54941/ahfe1003559

Measurement and Manipulation in Human-Agent Teams: A Review

Maartje Hidalgo[1], Summer Rebensky[1], Daniel Nguyen[1], Myke C. Cohen[1,2], Lauren Temple[1], and Brent Fegley[1]

[1]Aptima, Woburn, MA 01801, USA
[2]Arizona State University, Mesa, AZ 85212 USA

ABSTRACT

In this era of the Fourth Industrial Revolution, increasingly autonomous and intelligent artificial agents become more integrated into our daily lives. As such, these agents can conduct independent tasks within a teaming setting, while also becoming more socially invested in the team space. While ample human-teaming theories help understand, explain, and predict the outcome of team endeavors, such theories are not yet existent for human-agent teaming. Furthermore, the development and evaluations of agents are constantly evolving. As a result, many developers utilize their own test plans and their own measures making it difficult to compare findings across agent developers. Many agent developers looking to capture human-team behaviors may not sufficiently understand the benefits of specific team processes and the challenges of measuring these constructs. Ineffective team scenarios and measures could lead to unrepresentative training datasets, prolonged agent development timelines, and less effective agent predictions. With the appropriate measures and conditions, an agent would be able to determine deficits in team processes early enough to intervene during performance. This paper is a step in the direction toward the formulation of a theory of human-agent teaming, wherein we conducted a literature review of team processes that are measurable in order to predict team performance and outcomes. The frameworks presented leverage multiple teaming frameworks such as Marks et al.'s (2001) team process model, the IMOI model (Ilgen, 20005), Salas et al.'s big five model (2005) as well as more modern frameworks on human agent teaming such as Carter-Browne et al. (2021). Specific constructs and measures within the "input" and "process" stages of these models were pulled and then searched within the team's literature to find specific measurements of team processes. However, the measures are only half of the requirement for an effective team-testing scenario. Teams that are given unlimited amount of time should all complete a task, but only the most effective coordinative and communicative teams can do so in a time efficient manner. As a result, we also identified experimental manipulations that have shown to cause effects in team processes. This paper aims to present the measurement and manipulation frameworks developed under a DARPA effort along with the benefits and costs associated with each measurement and manipulation category.

Keywords: Human-agent teaming, Human systems integration, Artificial intelligence, Systems modeling language

INTRODUCTION

Intelligent agents are software agents with autonomous and intelligent capabilities (Russell & Norvig, 2009). These agents can be embodied (e.g., Boston Dynamics' ATLAS robot), or disembodied (e.g., Siri). Altogether, these capabilities allow artificial intelligence (AI) agents affordances like never before. In this era of the Fourth Industrial Revolution, technology is said to merge with humans (Schwab, 2017), and indeed, intelligent technologies like Siri or smart homes are commonplace. In this merge, agents may serve different roles. For example, AI can be a tool that is used by humans, a teammate that collaborates with us, or a decision aid or advisor to the human. The required capabilities of AI therefore differ given the context in relation to humans. Although this notion sounds common sense, the reality is that rigorous scientific understanding of human-agent teaming (HAT) is wanting, and agents are developed and implemented without the ability to understand and predict the effects. This paper is a step toward the formulation of a HAT theory with the development of frameworks around team processes. These frameworks scope the potential HAT-relevant constructs and manipulations, that could be measured by AI while predicting performance and outcomes, with a foundation in well-known models of teamwork.

MODELS OF TEAMWORK

Marks et al.'s Team Process Model

Team processes are central in the I-P-O models that dominated early theoretical models of teaming (e.g., McGrath, 1964). Marks et al. (2001, p. 357) defined team processes as "members' interdependent acts that convert inputs to outcomes through cognitive, verbal, and behavioral activities directed toward organizing taskwork to achieve collective goals." Through a review of the literature, the authors integrated the concept of *episodes* as temporal units of teaming (Mathieu & Button, 1992) into a recurring phase model that typifies such episodes as either action phases (in which teams engage in taskwork towards directly accomplishing team goals), or transition phases (in which teams reflect evaluate past action phases or plan for succeeding ones). The model identifies three major categories of team processes, depending on which phase they are more likely to occur: (1) *action processes* (i.e., progress monitoring, systems monitoring, team monitoring and backup behavior, and coordination); (2) *transition processes* (i.e., general mission analysis, formulation, planning, goal specification, and strategy formulation), and; (3) *interpersonal processes* (i.e., conflict management, motivation and confidence building, and affect management). Action and transition processes are more likely to occur in their corresponding phases, and interpersonal processes occur all throughout both types of phases.

Space constraints preclude an in-depth discussion of each category of team processes identified by Marks et al. (2001). Aside from their taxonomical framework, however, we note that this model was also among the first to distinguish team processes from team emergent states (i.e., the dynamic collective states of a team as they engage in team- and individual-level activities, such as cohesion and situation awareness). This distinction has since influenced the development of temporal and interaction-based measures and manipulations of emergent team cognitive states (e.g., Cooke et al., 2013) described in this paper.

IMOI Model

In combining McGrath's, (1964) I-P-O model of teamwork with Marks and colleagues' (2001) model of team processes, team processes can be understood as how a team takes inputs and convert them into outcomes. While this explains what a team does to produce their results, it does not capture the influence that interpersonal interactions have on the process itself and on the resulting outcomes. In addition to enacting team processes, teams also carry various cognitive, affective, and motivational states that exist and change as a result of team contexts and interactions (Marks et al., 2001; Rapp et al., 2021). These states, such as trust (Mayer et al., 1995), positive/negative affect (Tanghe et al., 2010), and cohesion (Carron et al., 2007) develop over time through interactions to emerge as a collective team state (Carter et al., 2015).

In 2005, Ilgen and colleagues adapted McGrath's, (1964) model to acknowledge how emergent states affect team functioning. Expanding the I-P-O model into the IMOI model, the term "mediators" was updated to represent the fact that both team processes and team emergent states play an important role in how a team converts their inputs into outputs.

Although the inclusion of emergent states in the model represents an important update in and of itself, the IMOI model elucidates two important concepts for team functioning. First, nesting team processes and team emergent states together under the "mediator" category implies that they reciprocally influence one another. Just as team processes give form to emergent states through team member interactions, emergent states also influence how team members interact. Second, the addition of the second "I" and the removal of the hyphenation between letters demonstrate that these team inputs, mediators, and outputs operate cyclically and non-linearly. Rather than seeing outputs as endpoints, they may also be considered as future inputs to future team functioning. How well a team performs at the end of one performance cycle can influence future strategies and interactions as well as how team members think and feel about one another. Ultimately, the IMOI model demonstrates that team functioning can be broken down into how team members interact (team processes) and how team members think and feel (team emergent states). These mediators affect performance, which in turn can affect future team processes and emergent states in a continuous cycle throughout a team's lifespan.

Salas et al.'s Big Five Model

Salas and colleagues (2005) conducted an extensive literature review to develop a scientifically sound yet practical model of teamwork. This model is distinct from the aforementioned models as it covers both constructs pertaining to teamwork processes (or the coordinating mechanisms of teamwork) and constructs of teamwork. They posited that effective teamwork requires all of the following coordinating mechanisms: (1) shared mental models or a shared understanding of goals, progress and action, (2) mutual trust which facilitates a willingness to rely on one another and cooperate together, and (3) closed-loop communication, wherein a message is sent, interpreted and acknowledged, and followed up on. Furthermore, these three constructs affect the core constructs of teamwork in often multiple ways. The core constructs include leadership, mutual performance monitoring, back-up behavior, adaptability, and team orientation. Although a full review of each of the constructs is too in-depth for the purpose of this paper, the main takeaway is that effective teamwork, according to the Big Five model, is a combination of the presence of core elements as well as reinforcing mechanisms. Human teamwork as such is a complex phenomenon that is not a mere emerging phenomenon, nor merely a process; teamwork is both.

Carter-Browne et al.'s Multidisciplinary Model of Teamwork for AI

Carter-Browne et al. (2021) focused on the previously discussed teaming models with human-agent or human-AI teaming constructs specifically under the context of agents assisting human to perform more effectively. Carter-Brown et al.'s multidisciplinary model of teamwork for AI specifically outlines the following elements at the organizational, team, human, and AI levels: (a) inputs of relevance, (b) processes and emergent states that could act as mediators and moderators, and (c) the outputs of each of these processes. Inputs can include cultural values, KSAOs, agent level of autonomy, and human personality. Processes and emergent states can include processes such as coordination, surveillance, AI explainability, and human cognitive load. Outputs can include organizational productivity, team effectiveness, AI error rates, and human efficiency. Additionally discussed by Carter-Browne et al. are considerations for elements such as task design, team type, interdependence, task uncertainty, job design, and temporal factors.

In summary, each of these models help explain and predict teamwork in different ways yet overarching conclusive notions about salient constructs and manipulations for studying effective HAT are lacking. In this next section, we discuss the specific methods used to develop construct and manipulation frameworks for HAT.

METHOD

Specific constructs and measures within the "input" and "process" stages of these models were pulled and searched on Google Scholar within the robotics, human-human and human-agent teams' literature to find specific measurements of team functions including inputs, communication, coordination, emergent states, resilience, and outputs resulting in a total of 110 teaming measurement and manipulation articles reviewed. We also identified experimental manipulations that have shown to cause effects in teaming measures. Additionally, forward and backward citation searching was conducted on the models referenced above. Over 150 measures and 130 manipulations were identified with a clear need to map them to higher level constructs and process areas. Using a multidisciplinary team, workshops were held to determine which measures and manipulations to condense, which ones were not feasibly captured by current HAT research capabilities, and the best construct and process mapping for each measure.

RESULTS

Measures

After identifying 150+ teams-related measures from the literature review, six categories were developed to organize these measures in a way that would reflect realistic team functioning. The organization of this construct framework loosely reflects the IMOI model (Ilgen et al., 2005) of teamwork, as it begins with inputs and ends with outputs that can subsequently serve as inputs. However, because mediator-type constructs (e.g., processes and emergent states) varied in the quantity of measures that have been used to capture them, four distinct categories were created for mediators rather than grouping all measures into one (e.g., "mediators") or two (e.g., team processes and emergent states) categories. These are summarized in Figure 1.

The first category of constructs, *Inputs*, refers to resources or existing influences that affect future team functioning. Input measures capture individual differences regarding the members themselves and are thought to influence how future team interactions may unfold. Such measures are beneficial when studying teamwork as they are either meaningful predictors of team mediators (e.g., skills/abilities) or are important to account for due to their significant influence on team mediators (e.g., propensity to trust). They are often captured through one-time surveys, making them cost-efficient. However, because so many input measures exist which are primarily subjective and collected via surveys, they have the potential to require subject participants to long periods of non-teaming tasks to collect the data necessary, in addition to requiring an ability for the AI to ingest the survey data before performing in order to reason over teaming tasks.

The second two categories of constructs, Communication and Coordination, refer to team processes that teams enact to complete tasks (Marks et al., 2001). These categories were explicitly separated as many different measures were observed for each of these processes. The *Communication* category encompasses all measures related to how (communication content) and when (temporal communication) team members convey and receive information, and the distribution of information (communication pattern). Communication measures are beneficial in that they are abundantly available in many forms and can be highly informative for understanding nuances of teamwork (Cooke et al., 2013). However, they are difficult to analyze due to the diverse types of measures, natural language processing requirements, and clear theoretical guidance needed to interpret communication measures. Communication-related measures are currently complex, time- or resource-intensive, and infeasible to process for many AI agents.

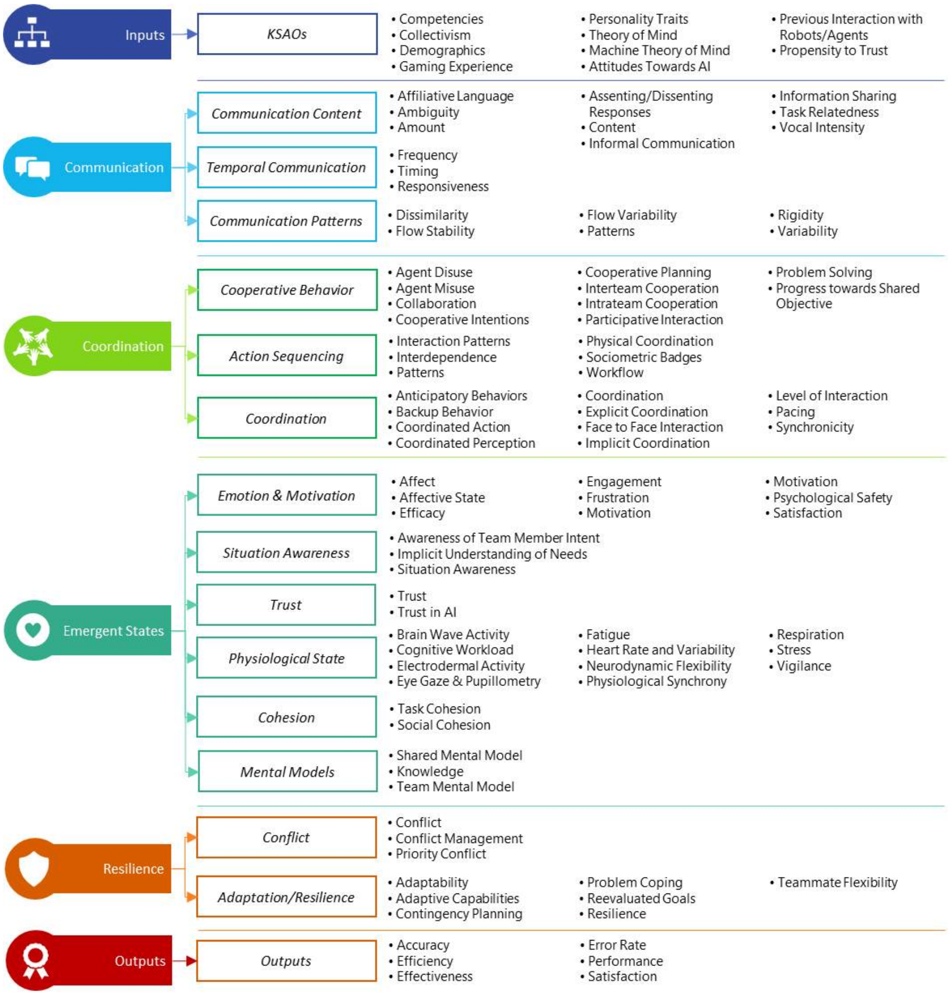

Figure 1: Constructs framework for human-agent teaming.

The *Coordination* category encompasses all constructs related to the nature of how team members choose to work together (cooperative behavior), the patterning and flow of joint actions by team members (action sequencing), and taskwork planning (coordination). Measures of coordination are highly valuable for understanding exactly how teams convert inputs into outputs and can provide useful datapoints above and beyond serving as measures of teamwork (e.g., identifying points of intervention). However, like communication measures, they can be difficult to analyze and require meaningful forms of aggregation that are guided by theory. Team coordination can be measured in simulation-based settings as each team members movement and interactions can be tracked; however, methods for real-time measurements of team coordination are limited in practice (cf., Gorman et al., 2012).

The *Emergent States* category encompasses all constructs that relate to non-behavioral team experiences such as cognitions (e.g., mental models), attitudes (e.g., trust), affect (e.g., emotions), and physical responses (e.g., physiological states). Emergent states are an important part of team functioning and can be thought of as the psychological underpinnings that both drives and is driven by team processes (Ilgen et al., 2005). The benefits of these measures are that they thus elucidate key components of team functioning that are not captured by observable process measures. However, they take time to unfold (Carter et al., 2015), emerge in various forms depending on the emergent state (Rapp et al., 2021), and can be difficult to parse for the less apparent emergent states such as trust and emotion. Additionally, some emergent states can be costly to measure, such as physiological states which require sensors to capture. Leveraging subjective measures, behavioral indicators, and physiological metrics provide a rich source of in-the-moment fluctuations in team processes (see Figure 1).

The *Resilience* category encompasses constructs that relate to how teams withstand adversity. This category is unique in that the constructs within it (conflict and adaptation) can refer to both team processes (e.g., conflict management, contingency planning) or emergent states (e.g., conflict states, adaptive capacity). Resilience measures thus provide deeper insights into how teams adaptively react to adverse situations as a function of team processes and emergent states. However, this also means that care must be taken to not conflate these measures. For example, a measure of a team's conflict management strategies is not suitable for inferring their average level of conflict. Additionally, depending on the team and setting, conflict may be few and far between, which could complicate the ability to train agents under conflict settings.

Lastly, the *Outputs* category encompasses constructs regarding team performance with respect to team goals. Although any team mediator could be considered an output depending on the interest at hand, this framework views outputs as the product of teamwork. That is, measures that are influenced by team functioning, but not an actual part of team functioning it and of itself. Measuring an output is beneficial in that it is often the ultimate outcome of interest. One of the costs of measuring outputs is that there are both subjective (e.g., ratings) and objective (e.g., accuracy scores) measures of performance, which can differ in their relationships to measures of team mediators. Ensuring a comprehensive picture of what "performance" means for human-agent team settings will ensure agents can balance and prioritize the many beneficial outcomes of teams.

Manipulations

Our review yielded three major categories of manipulations (Figure 2): agent capabilities and characteristics, task, and team. A fourth category, environmental context, is excluded due to space constraints.

The manipulation of agents in the current HAT literature primarily focuses on the effects of various AI capabilities on teaming processes and emergent states. These include AI processing efficiencies and reliability manipulations to affect team performance outcomes (e.g., Appelganc et al., 2022), as well as explanation and affect management to dampen or amplify the effects of

Figure 2: Manipulation framework for human-agent teaming.

perceptive team emergent states and attitudes like trust on team performance (e.g., Textor et al., 2022). In addition, several studies manipulate agent characteristics like anthropomorphism and embodiment towards directly investigating their effects on people's perceptions of AI teammates (de Visser et al., 2022). Manipulating agent algorithms arguably generates the most direct causal links between AI design frameworks and a broad range of HAT measures. However, the development of novel AI features may require advanced computer science expertise beyond the capacity of many human factors research units. Cooke et al. (2020) outlines agent manipulation methods that eschew actual AI development through Wizard-of-Oz frameworks, in which highly trained confederates simulate theoretical AI teammate features but are portrayed as synthetic teammates to participants.

Task manipulations center on observing how hybrid teams are affected by differences in taskwork, i.e., team members' interactions with tasks, technology, and systems to fulfill individual-level responsibilities (Bowers et al., 1997). These abound as manipulations of task fulfillment requirements, such as time and resource constraints (Chiou & Lee, 2016) and exclusivity of task types to certain roles or time periods (Freeman et al., 2022). Notably, some studies instead manipulate the availability of supplementary information and resources to ease individual task completion from technological aids, the environment, or the task structure itself (e.g., Corral et al., 2021). Individual manipulations can be applied at various levels of team membership and at different timescales with minimal changes to testbeds, task environments, or administrative protocols. However, we also observed that task manipulations are often implemented as a combination of several taskwork attributes at a time. The cost of task manipulations, therefore, is dependent on both type *and* quantity of taskwork attributes being manipulated.

Manipulations of team cognitive attributes and processes primarily affect how team interactions take place, including communication interfaces, modalities, and technologies made available to the team. Many team manipulations additionally alter team structural contexts that govern interactions, such as through the specification of role hierarchies, responsibilities, and interdependencies between team members. In some cases (e.g., Gorman et al., 2010), the individuals that comprise a team are also shuffled; variations of this include manipulations of spatiotemporal distribution. We note that direct manipulations of team interaction media can entail designing several variants of HAT testbeds, which can be limiting for complex teaming designs. It may be more cost-effective—albeit also limited—to manipulate team goals vis-à-vis individual goals. This may capture how the limited capabilities of AI to prioritize team-level goals over individual ones (Chiou & Lee, 2016) affect overall teaming in larger HATs.

Finally, we note that momentary perturbations can be applied to the majority of variables listed in Figure 2. Perturbations are especially relevant for temporal questions about human-agent teams. We refer the reader to Cooke et al. (2020) and Gorman et al. (2017) for an overview of perturbation-based manipulation designs and analyses, many of which notably incur minimal implementation costs.

DISCUSSION

With the surge of agent development and the fusion of agents into human teaming situations, researchers and developers need to gain a better understanding of the components that affect teamwork. Questions such as, in what way will the agent impact the mission or team dynamic, remain to be answered. This paper is a step toward a more comprehensive understanding of human-agent teaming (HAT). To that end, we reviewed the literature to extract measures and manipulations of HAT and consolidated these components into a reusable framework. Although this framework is not exhaustive, it does focus on teaming functions that can be observed, measured, and leveraged by AI in a collaborative or predictive capacity and are prevalent indicators of teaming success across the areas of team and AI research. Additionally, the presented manipulations provide specific elements that can be altered in environments such as simulation, to develop AI that can consider the factors of the team and task in its collaboration with, prediction or assessment of teams. Future work is aimed at contextualizing the frameworks and determining preferences to use by AI due to their impact on team performance and ease/cost of implementation. We intend to identify manipulations that have the greatest impact on each construct category to determine the best testing conditions for AI in human-agent teams, while additionally highlighting gaps that leave room for innovation. The results of these efforts aim to create an overarching theory of human-AI teams.

ACKNOWLEDGMENT

The research reported in this document/presentation was performed in connection with contract number W912CG-22-C-0001 with the U.S. Army Contracting Command - Aberdeen Proving Ground (ACC-APG) and the Defense Advanced Research Projects Agency (DARPA). The views and conclusions contained in this document/presentation are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of ACC-APG, DARPA, or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. Distribution Statement "A" (Approved for Public Release, Distribution Unlimited).

REFERENCES

Appelganc, K., Rieger, T., Roesler, E., & Manzey, D. (2022). How Much Reliability Is Enough? A Context-Specific View on Human Interaction With (Artificial) Agents From Different Perspectives. *Journal of Cognitive Engineering and Decision Making*, 155534342211046. https://doi.org/10.1177/15553434221104615

Bowers, C. A., Braun, C. C., & Morgan, B. Jr. B. (1997). Team Workload: Its Meaning and Measurement. In *Team Performance Assessment and Measurement*. Psychology Press.

Carter-Browne, B. M., Paletz, S. B. F., Campbell, S. G., Vahlkamp, S. H., Schwartz, J., & O'Rourke, P. (2021). There is No "AI" in Teams: A Multidisciplinary Framework for AIs to Work in Human Teams. 94.

Carter, N. T., Carter, D. R., & DeChurch, L. A. (2018). Implications of observability for the theory and measurement of emergent team phenomena. *Journal of Management*, *44*(4), 1398–1425.

Carron, A. V., Eys, M. A., & Burke, S. M. (2007). Team cohesion: nature, correlates, and development. *Social Psychology in Sport*, 91-101.

Chiou, E. K., & Lee, J. D. (2016). Cooperation in Human-Agent Systems to Support Resilience: A Microworld Experiment. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(6), 846–863. https://doi.org/10.1177/0018720816649094

Cooke, N., Demir, M., & Huang, L. (2020). A Framework for Human-Autonomy Team Research. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics. Cognition and Design* (pp. 134–146). Springer International Publishing. https://doi.org/10.1007/978-3-030-49183-3_11

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science*, *37*(2), 255–285. https://doi.org/10.1111/cogs.12009

Corral, C. C., Tatapudi, K. S., Buchanan, V., Huang, L., & Cooke, N. J. (2021). Building a Synthetic Task Environment to Support Artificial Social Intelligence Research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 660–664. https://doi.org/10.1177/1071181321651354a

de Visser, E. J., Topoglu, Y., Joshi, S., Krueger, F., Phillips, E., Gratch, J., Tossell, C. C., & Ayaz, H. (2022). Designing Man's New Best Friend: Enhancing Human-Robot Dog Interaction through Dog-Like Framing and Appearance. *Sensors*, *22*(3), Article 3. https://doi.org/10.3390/s22031287

Freeman, J., Huang, L., Wood, M., & Cauffman, S. J. (2022). Evaluating Artificial Social Intelligence in an Urban Search and Rescue Task Environment. In N. Gurney & G. Sukthankar (Eds.), *Computational Theory of Mind for Human-Machine Teams* (pp. 72–84). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-21671-8_5

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010). Training Adaptive Teams. Human Factors, 52(2), 295–307. https://doi.org/10.1177/0018720810371689

Gorman, J. C., Dunbar, T. A., Grimm, D., & Gipson, C. L. (2017). Understanding and Modeling Teams As Dynamical Systems. *Frontiers in Psychology*, *8*, 1053. https://doi.org/10.3389/fpsyg.2017.01053

Gorman, J. C., Hessler, E. E., Amazeen, P. G., Cooke, N. J., & Shope, S. M. (2012). Dynamical analysis in real time: Detecting perturbations to team communication. *Ergonomics*, *55*(8), 825–839. https://doi.org/10.1080/00140139.2012.679317

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. Annual Review of Psychology, 56, 517–543.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. Academy of Management Review, 26(3), 356–376.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of management review, 20(3), 709-734.

McGrath, J. (1964). Social Psychology: A Brief Introduction. New York: Holt, Rinehart and Winston.

Rapp, T., Maynard, T., Domingo, M., & Klock, E. (2021). Team emergent states: What has emerged in the literature over 20 years. *Small Group Research*, *52*(1), 68–102.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555–599. https://doi.org/10.1177/1046496405277134

Schwab, K. (2017). *The Fourth Industrial Revolution*. Crown.

Tanghe, J., Wisse, B., & Van Der Flier, H. (2010). The formation of group affect and team effectiveness: the moderating role of identification. *British Journal of Management*, *21*(2), 340-358.

Textor, C., Zhang, R., Lopez, J., Schelble, B. G., McNeese, N. J., Freeman, G., Pak, R., Tossell, C., & de Visser, E. J. (2022). Exploring the Relationship Between Ethics and Trust in Human–Artificial Intelligence Teaming: A Mixed Methods Approach. *Journal of Cognitive Engineering and Decision Making*, *16*(4), 252–281. https://doi.org/10.1177/15553434221113964