SUBMISSION TYPE
Poster

TITLE
Measuring Multiteam System Performance with Multi-Objective Optimization

ABSTRACT
We propose an advanced multi-objective optimization approach to stratify and measure performance of individuals, teams, and multiteam systems with competing local objectives, a shared global objective, and complex task interdependence. Novel measures were empirically tested on a timed cooperative activity and validated against traditional statistics-based performance measures. To assess the external validity of these metrics, structural network features of the multi-team systems were found to be more correlated with our new measures of performance than traditional performance measures.

WORD COUNT
2864 words

Introduction

Why are some teams more successful than others? Decades of research have focused on understanding performance of individuals, teams, and more recently, multiteam systems (MTS), with the goal of generating research, providing teams with feedback, developing team training, evaluating performance, and planning for the future (Wildman et al., 2012). However, there has been limited development on the tools used for measuring performance. By developing better tools to measure performance, we can accomplish a deeper understanding of individual and social determinants that promote success in individuals, teams, and MTS.

Multiteam systems describe networks of interdependent teams (Zaccaro et al., 2020). The extant literature conceptualizes MTS performance as the success to which interdependent teams achieve a collective goal (de Vries et al., 2016; Marks et al., 2005; Murase et al., 2014). While component team performance has been studied alongside MTS performance (DeChurch & Marks, 2006), there has not been a focus on understanding individual, team, and MTS performance together. What if, despite having a collective goal, the MTS component teams or individuals have conflicting objectives? An ideal decision for one individual may lead to a poor outcome for the MTS as a whole. Our goal is to determine an effective methodology to capture performance not only within each level of an MTS, but also across each level. For example, how well did an individual complete not only their individual objectives but also to what extent did they contribute to the team or multiteam system objectives?

Regardless of which level performance is studied, researchers have reached a consensus on its conceptualization: performance captures the extent to which a party accomplishes an established objective. However, it can be very tricky to actually measure this concept. Operationalization of performance measurement occurs in the following ways: statistics (ex. number of targets reached, objectives completed, total cost), human raters (ex. peer evaluation, subject-matter expert panel, manager evaluation), and unobtrusive measurement (ex. sensors, emails, messages) (Lanaj et al., 2018; Pilny et al., 2014). Each method of operationalizing performance measurement experiences its own limitations. Statistics poorly capture the multilayered and interdependent environment in which individuals, teams, and MTS are embedded. Human raters are subject to bias, and the active presence of external observers can alter the behavior of the subjects being measured (Kingstrom & Mainstone, 1985). Furthermore, measuring performance requires time and attention from trained experts, which can be costly. While unobtrusive measurement techniques will not alter the behavior of subjects, they are frequently not explainable. Thus results may not be perceived as trustworthy. Most importantly, none of these measures capture the performance across layers of a multi-team system.

In this study, we develop a novel approach to measuring performance in multilevel environments by utilizing a multi-objective optimization approach. The developed measures are unbiased, nuanced, unobtrusive, and explainable. Furthermore, they incorporate individual and team conflict and dependencies to present detailed measures describing the inter-level performance of individuals and teams in MTS. We present the analytic approach used, an empirical application, and evidence of construct validity.

Performance of Individuals, Teams, and MTS: Teams are composed of two or more individuals interacting with meaningful interdependencies, shared goals, and embedded in a context influenced by ongoing processes (Salas et al., 2007). A multiteam system (MTS) comprises "two or more teams that interface directly and interdependently in response to environmental contingencies toward the accomplishment of collective goals" (Mathieu et al. 2001). In simpler words, MTS are teams of teams. How does performance measurement vary across individual, team, and MTS scales? Measuring individual performance typically involves self, peer, or supervisor evaluation. Teams can be measured in aggregate, using external raters, or based on statistical measurements of team-based outcomes. Alternatively, team performance can be measured by averaging across individuals. MTS performance is frequently operationalized by capturing completion rate of specific tasks or objectives (DeChurch & Marks, 2006; Pilny et al., 2014). These approaches do not include methods of capturing inter-level performance (i.e. direct contributions of an individual to the team and/or MTS global objective) as well as a method to capture interdependencies within each level. Our goal is to propose a framework which includes both of these features.

Framework

The proposed framework begins with model definition, proceeds by solving key parameters, and concludes with deriving a performance measure. Model definition includes specifying the measurement targets through system decomposition, determining objectives for each component in the system, and developing a computational model to describe the system. Key parameters incorporating interdependencies are solved as follows: 1) determine bounds for scaling, 2) compute the best and worst feasible decisions, and 3) calibrate the observed decisions. Finally, a performance measure is derived by comparing the calibrated observed decision to the best and worst feasible decisions.

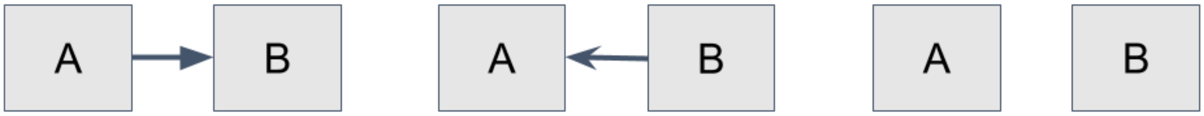
Model Definition: This framework is designed for capturing interdependencies within and across teams in MTSs collaborating on a shared objective. Thus, an MTS should be decomposed hierarchically into smaller components. The basic framework is explained with a system of individuals, teams, and a single MTS, but the proposed framework can be extended to multiple MTSs working interdependently. For a system of individuals, teams, and a single MTS, measures will be computed within-level: individual-oriented individual, team-oriented team, and MTS-oriented MTS performance. Additionally, **performance is computed across-levels**: team-oriented individual, MTS-oriented individual, and MTS-oriented team performance. By computing across levels, it is possible to determine whether an individual may contribute to a positive team outcome, despite having poor individual performance. Conversely, an individual who optimizes their own individually-oriented performance to the detriment of the team can be identified through this framework.

Objective specifications can be unique for different individuals or teams, but they must be specific, measurable indicators with a goal of minimization or maximization. For example, the goal of achieving a balanced diet is unspecifiable. However, this objective can be reconstructed as minimizing the differences between a planned diet and the actual food consumed by an individual. Thus, it is possible to measure a wide range of objectives.

To develop a solvable computational model, approaches from mathematical optimization can be employed. Once appropriate objectives have been specified for each individual, every team, and the MTS, we define an objective function that is minimized or maximized based on the objective specification. All possible decision alternatives are specified as variables. Then, environmental constraints are modeled to constrain the objective function. A complete model should have exactly one objective function and can have multiple variables and constraints for each individual, each team, and the MTS.

Solving Key Parameters: After a model has been specified, the next step is to determine key values that minimize or maximize the objective and incorporate interdependencies. We begin by observing the dependence structures and will follow with a detailed discussion of the parameter calculation incorporating interdependencies. There are three types of dependencies:

1) A leads B
2) A follows B
3) Pooled interdependence



Type 1: A leads B

Type 2: A follows B

Type 3: Pooled interdependence

To account for these types of interdependence, the actual decision can be calibrated by simulating optimal decisions of other individuals or teams who may affect the target's performance. A type 1 dependency describes an individual A following individual B. Suppose we observe two individuals working together to make a pizza: Ayman (A) and Bea (B). Their team objective is to produce the best pizza. Ayman prepares the pizza dough and toppings and Bea bakes the pizza in the oven after the pizza is prepared. Thus, A and B are interdependent, where A leads B, and thus B follows A. Both Ayman and Bea will impact each others' team-oriented performance.

To appropriately compute a team-oriented individual score for Ayman, we follow an approach for a type 1 dependency: A leads B. For example, if Bea burned the pizza in the oven, Ayman should not be penalized if he prepared the pizza perfectly. When a type 1 dependency occurs, the observed decision is retained, with comparison points being adjusted to incorporate the impact of B's decisions on A's team-oriented individual performance score.

To compute a team-oriented individual score for Bea, we follow the approach for a type 2 dependency: B follows A. With a type 2 dependency, the actual observed decision made by B must be adjusted based on the decision made from A. So, if Ayman forgets to put any toppings on the pizza, but Bea cooks the pizza perfectly, Bea should not be penalized for Ayman's mistake. Thus, the final outcome is adjusted by simulating A performing optimally.

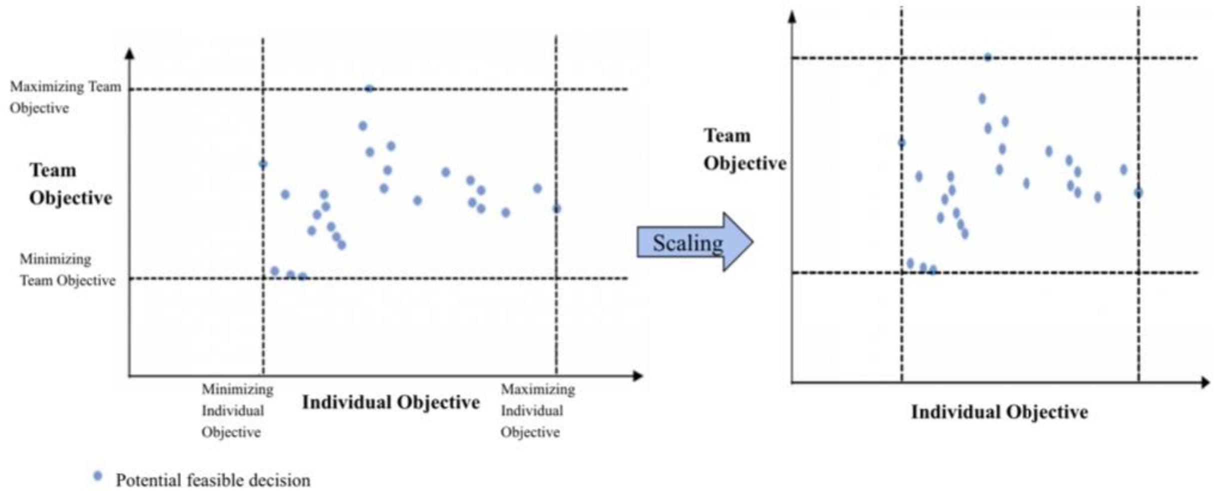
The final form of dependence is pooled interdependence (Bishop & Mahajan, 2005). While A and B may not directly impact each others' decisions, they are interdependent in the sense that they collaboratively contribute to a system objective. Because they do not impact each other's decisions directly, there is no calibration or adjustment needed to proceed with the performance calculation.

We will begin by computing team-oriented individual performance as a descriptive example. Suppose the individual and the team objectives both should be maximized. We will compute upper and lower bounds for scaling the objectives, obtain the best and worst feasible decisions, and calibrate the observed decision based on dependencies on other individuals or teams.

1) Determine bounds for scaling:

The first step is to compute bounds of each objective function independently with the goal of normalizing individual and team objectives. Thus, the individual objective is maximized without being constrained by the team decisions. A minimum, or worst possible, decision is also computed to provide a lower bound. After the individual objective bounds have been computed, the process is repeated for team objectives. Then, the team objective can be rescaled by a scaling factor.

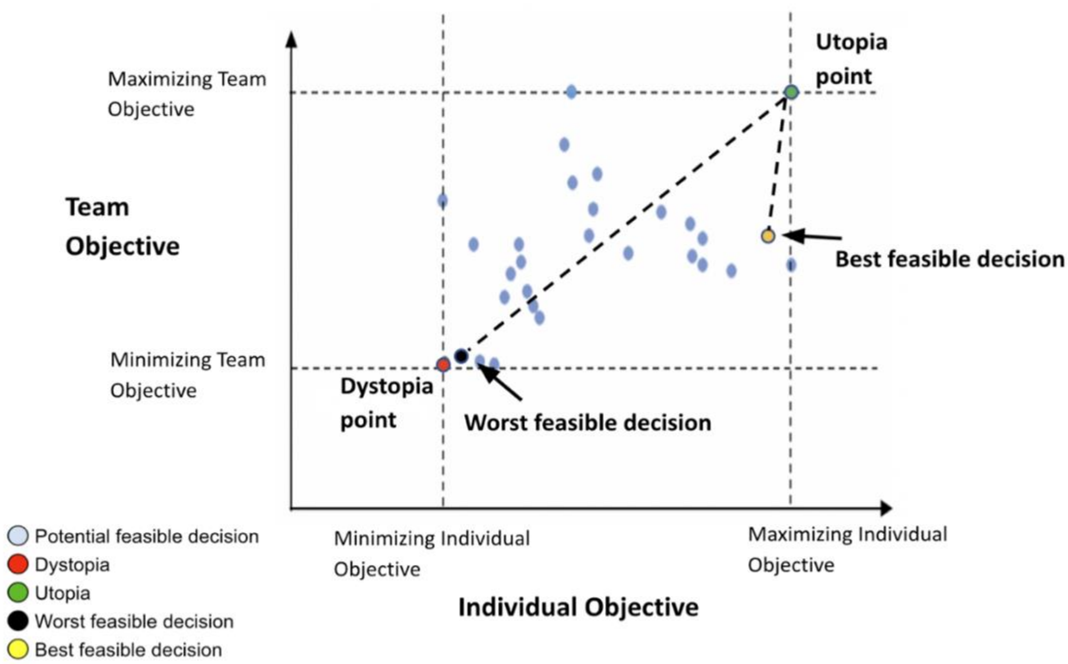
$$\text{Scaling Factor(SF)} = \frac{\text{Individual Objective Upper Bound} - \text{Individual Objective Lower Bound}}{\text{Team Objective Upper Bound} - \text{Team Objective Lower Bound}}$$



Incorporating interdependence:
If a type 1 dependence (A leads B) exists, then the individual and team objective bounds must be adjusted by fixing decisions of the follower B. Thus the ideal individual and team decisions obtained in the model will incorporate the dependence structure, and the team-oriented individual performance rating of A will be scaled based on the decisions already made by B. With this adjustment, A will not be penalized for poor performance of B. A type 2 (A follows B) or type 3 (null) dependence structure requires no adjustment to the bounding process.

2) Compute the best and worst feasible decisions:

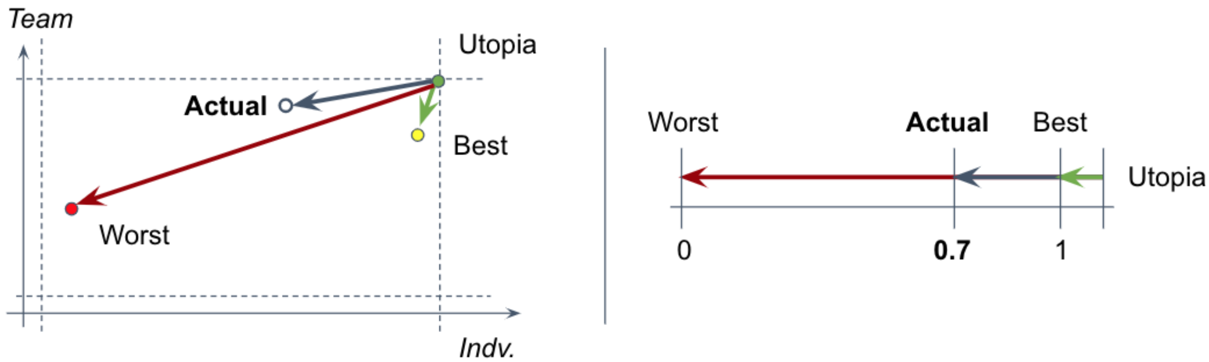
After the objectives are rescaled, the intersection of the bounds that maximizes both objectives creates an optimal "utopia" point, which is generally infeasible to obtain. Thus, the best feasible decision can be obtained by finding the feasible decision that minimizes the Euclidean distance to the utopia. Similarly, the worst feasible decision represents the feasible decision that maximizes the distance to the utopia point.



Incorporating interdependence:
If a type 1 dependence (A leads B) exists, then the best and worst feasible points must be adjusted by fixing decisions of the follower B. Thus the ideal individual and team decisions obtained in the model will incorporate the dependence structure, and the team-oriented individual performance rating of A will be scaled based on the decisions already made by B. With this adjustment, A will not be penalized for poor performance of B. A type 2 (A follows B) or type 3 (null) dependence structure requires no adjustment to the calculation of best and worst feasible points

3) Calibrate the observed decision:
Once comparative measures have been obtained, the next step is to calibrate the actual decision based on interdependence with other components of the MTS. This step is only required for a type 2 dependence (A follows B). In this step, we observe the decision made by individual A and calibrate their team-oriented outcome by optimizing the decision made by individual B.

Computing a performance measure:

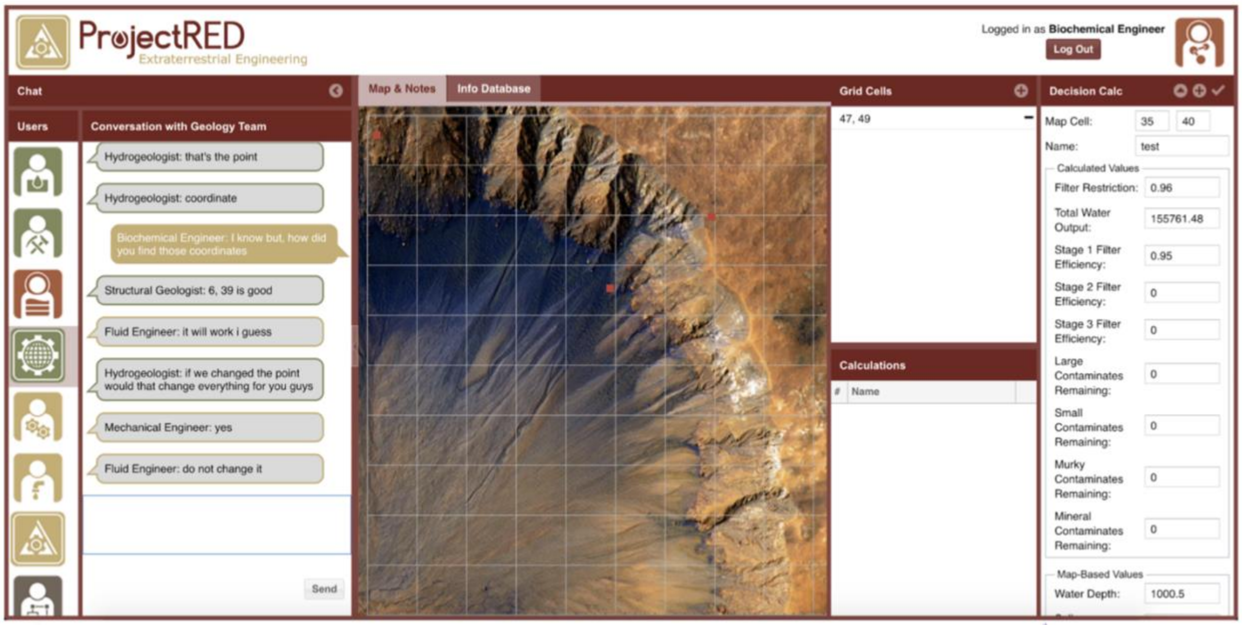


After establishing appropriate comparisons and adjusting the observed value as necessary, a performance measure can be computed. An optimal decision matches the best feasible decision, which is the feasible decision which minimizes the Euclidean distance between the decision and the utopia point. Thus, if a decision's distance to the utopia point matches the distance of the best feasible point to utopia, then the decision should receive a perfect performance rating. Similarly, if the decision's distance to the utopia point matches the distance of the worst feasible point to utopia, it should receive a performance rating of zero, or the worst possible score. Then, for any point in between, the score can be computed by normalizing the distance of the calibrated observed decision to utopia based on the difference between the best and worst feasible distances from utopia. This is formalized by computing a "lack of performance" measure and subtracting its value from 1 to obtain a final performance score.

$$\mathrm{LP}_{\mathrm{i}\in\mathcal{S}} = \frac{\|(\mathrm{Calibrated\ Observation})_{\mathrm{i}} - \mathrm{Utopia}_{\mathrm{i}}\| - \|(\mathrm{Best\ Feasible})_{\mathrm{i}} - \mathrm{Utopia}_{\mathrm{i}}\|}{\|(\mathrm{Worst\ Feasible})_{\mathrm{i}} - \mathrm{Utopia}_{\mathrm{i}}\| - \|(\mathrm{Best\ Feasible})_{\mathrm{i}} - \mathrm{Utopia}_{\mathrm{i}}\|}$$

Application

This theoretical framework was empirically tested on a design team activity Project RED (Red planet, Exploration and Development) Design. Participants in Project RED Design are tasked to provide sustainable, clean water to future inhabitants of Mars. The task involves an MTS comprising four 3-member teams with various expertise in the design of water infrastructure: planetary geology, extraterrestrial engineering, space robotics, and space human factors. The overarching objectives of the Project RED Design MTS are to maximize water yield, water quality, and accessibility. However, each team is also optimizing their proximal goals in terms of water content in soil (planetary geology), cost of drilling equipment (extraterrestrial engineering), energy requirements to maneuver robots (space robotics), and proximity of water to human habitat (space human factors). Hence, the goal of the activity is to collectively design the best location and specification to maximize the supportable population while minimizing overall costs. Each individual has a unique role with different information available and the ability to control specific equipment parameters.



Participants use a computer interface to collectively identify an optimal decision, which allows for the computation of all feasible decisions. Computational models for each performance metric were developed using the AMPL optimization software, using the BARON global optimization solver. Novel metrics were then compared to existing performance metrics, which were computed by comparing performance across multiple replications of the same task.

Results

Novel performance measures were compared with traditional performance metrics as well as related measures to establish construct validity.

Convergent Validity: To validate the convergent validity of measures, correlations were computed between existing and novel measures of performance on two distinct datasets, named NEK 1 and HERA. Based on the results, novel metrics computed for individuals (individual, team, and MTS oriented individual performance scores) were more highly correlated with measures of individual performance than metrics for team and MTS performance (team and MTS oriented team and MTS oriented MTS). Similarly, novel team-related performance metrics (team and MTS oriented team) were most correlated with team performance. Finally, MTS performance was most correlated with MTS-oriented MTS performance.

	NEK 1			**HERA**		
	Indv.	**Team**	**MTS**	**Indv.**	**Team**	**MTS**
Indv. Oriented Indv.	**0.52*****	0.31**	0.21*	**0.76*****	0.77***	0.33

Team Oriented Indv.	0.40***	0.41***	0.24*	0.72***	0.73***	0.47^{+}
MTS Oriented Indv.	0.32**	0.45***	0.28**	0.68***	0.67***	0.59*
Team Oriented Team	0.21*	**0.58*** **	0.38***	0.56***	**0.72*** **	0.52*
MTS Oriented Team	0.28**	0.61***	0.36***	0.48***	0.67***	0.52*
MTS Oriented MTS	-0.02	0.07	**0.43*** **	0.56*	0.43^{+}	**0.88*** **

$^{+}p<0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Postdictive Validity: To validate the external validity of measures, social networks were constructed from survey data to determine whether novel measures were more correlated with external predictors of performance than traditional measures. The findings showed that team-oriented and MTS-oriented individual metrics were more significantly correlated with external measures compared to traditional measures of performance.

Question	Trad.	IoI	ToI	MoI
Who do you rely on for leadership?	0.033	0.015	**0.194** **	**0.201** **
Who was a valuable source of information?	0.112	0.134^{+}	0.200	**0.248*** **
With whom did you work effectively?	0.177*	0.225**	0.291***	**0.305*** **
Who was instrumental in helping the task force achieve its goals?	0.090	0.132^{+}	**0.231*** **	**0.245*** **

$^{+}p<0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

measures in **bold** are significantly different from the traditional measure by the Pearson-Filon test for overlapping correlations on dependent groups at a significance level of $p < 0.05$

Discussion

The framework proposed in this study provides a holistic approach to performance measurement, providing measures for performance within a hierarchical structure. By assessing performance on more axes, it may be possible to determine more specific mechanisms or drivers occurring beneath the surface. Furthermore, this performance measurement approach enables users to identify strong component members in weak teams or MTS (or vice versa). Additionally, this method doesn't require comparison to other replications, which makes it a strong choice for performance measurement of unique, dynamic, and cooperative activities or tasks.

The approach is subject to multiple limitations, which can be potential opportunities for future research. First, the framework is designed for cooperative tasks with a shared superordinate goal. Further research is required to extend the framework for competitive tasks. Additionally, this approach only provides a single dimension for each level of performance. This is acceptable for simple, straightforward tasks or activities where objectives are clearly defined, but it requires extensions to be adapted to environments where individuals may have multiple goals; i.e. a salesman may want to maximize his profit while minimizing the time or effort he puts into work. Finally, this approach provides a point-in-time computation of performance based on results of a single activity or event; the next frontier of performance quantification lies in capturing performance dynamically, with time as an equally important dimension in addition to structural level.

References

Bishop, J. W., & Mahajan, A. (2005). The Use of Teams in Organizations: When a Good Idea Isn't and When a Good Idea Goes Bad. *Laboratory Medicine*, *36*(5), 281–285.

DeChurch, L. A., & Marks, M. A. (2006). Leadership in multiteam systems. *The Journal of Applied Psychology*, *91*(2), 311–329.

de Vries, T. A., Hollenbeck, J. R., Davison, R. B., Walter, F., & van der Vegt, G. S. (2016). Managing coordination in multiteam systems: Integrating micro and macro perspectives. *Academy of Management Journal. Academy of Management*, *59*(5), 1823–1844.

Kingstrom, P. O., & Mainstone, L. E. (1985). An Investigation of the Rater-Ratee Acquaintance and Rater Bias. *Academy of Management Journal*, *28*(3), 641–653.

Lanaj, K., Foulk, T. A., & Hollenbeck, J. R. (2018). The Benefits of Not Seeing Eye to Eye with Leadership: Divergence in Risk Preferences Impacts Multiteam System Behavior and Performance. *Academy of Management Journal*, *61*(4), 1554–1582.

Marks, M. A., DeChurch, L. A., Mathieu, J. E., Panzer, F. J., & Alonso, A. (2005). Teamwork in multiteam systems. *The Journal of Applied Psychology*, *90*(5), 964–971.

Mathieu, J.E., Marks, M.A., Zaccaro, S.J.. (2001). Multiteam systems. *Organizational Psychology, Vol. 2: Handbook of Industrial, Work and Organizational Psychology*, 289–313.

Murase, T., Carter, D. R., DeChurch, L. A., & Marks, M. A. (2014). Mind the gap: The role of leadership in multiteam system collective cognition. *The Leadership Quarterly*, *25*(5), 972–986.

Pilny, A., Yahja, A., Poole, M. S., & Dobosh, M. (2014). A Dynamic Social Network Experiment with Multi-team Systems. *2014 IEEE Fourth International Conference on Big Data and Cloud Computing*, 587–593.

Salas, E., Stagl, K. C., Burke, C. S., & Goodwin, G. F. (2007). Fostering team effectiveness in organizations: toward an integrative theoretical framework. *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation*, *52*, 185–243.

Wildman, J. L., Thayer, A. L., Pavlas, D., Salas, E., Stewart, J. E., & Howse, W. R. (2012). Team knowledge research: emerging trends and critical needs. *Human Factors*, *54*(1), 84–111.
Zaccaro, S. J., Dubrow, S., Torres, E. M., & Campbell, L. N. P. (2020). Multiteam Systems: An Integrated Review and Comparison of Different Forms. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*(1), 479–503.