6 Computational Modeling of Long-Distance Space Exploration

A Guide to Predictive and Prescriptive Approaches to the Dynamics of Team Composition

Brennan Antone and Alina Lungeanu
Northwestern University

Suzanne T. Bell
KBR/NASA Johnson Space Center

Leslie A. DeChurch and Noshir Contractor
Northwestern University

CONTENTS

9781138339866_C006.indd 107 13/07/20 4:34 PM

INTRODUCTION

Over the centuries, humankind has taken on many challenging explorations that require collaboration because their very survival relied on it. Humanity has been collectively exploring beginning with the agrarian and nomadic ages, followed by maritime explorations climaxing with the discovery of the 'new world,' scaling the peaks of our tallest mountains, diving to the deepest trenches of our oceans and standing-up bold polar expeditions. Finding individuals to engage in these daredevil adventures is not for the faint of heart – and spirit. An observation that Sir Ernest Henry Shackleton, a British Antarctic explorer who led three expeditions to the Antarctic, was acutely aware. Some sources recount that when Sir Shackleton tried to recruit a crew for one of his Antarctic expeditions, his classified ad in the newspaper reportedly read: 'Men wanted for Hazardous Journey. Small wages, bitter cold, long months of complete darkness, constant danger, safe return doubtful. Honor and recognition in case of success' (Huntford, 2013). This ad was not intended to appeal to the legendary glamorous swashbuckling sailors immortalized in fiction. Indeed, in a study of 25 personnel who spent the 9-month austral winter confined to two small, isolated research stations on the Antarctic ice cap, Biersner and Hogan (1984) found that the most positive peer nominations were received by those who scored *low* on self-reflection and emotional expressiveness. Relatedly, based on several studies of human responses to life at the US Amundsen-Scott South Pole station, Natani and Shurley (1974, p. 90) concluded that the Antarctic station had become 'a haven for the technically competent individual who is deficient in social skills.'

It is within this much longer-term context that we must consider humanity's 20th-century foray into space. It is but the latest 'giant leap' that is building on an arguably equally significant arc of achievements by our ancestors. Having explored and exploited most of the frontiers on Earth, space travel puts us on the brink of making humans an interplanetary species. The public's interest in the rugged individualistic qualities that epitomized the very first astronauts in space – *the Right Stuff* – was captured in Tom Wolfe's 1979 eponymous book. Tom Wolfe focused on the qualities of the Mercury Seven – Scott Carpenter, Gordon Cooper, John Glenn, Gus Grissom, Wally Schirra, Alan Shepard, and Deke Slayton – who were all part of the first (and last) solo Mercury missions into space. As we progressed through subsequent space programs – Gemini, Apollo, Skylab, Space Shuttle, and the International Space Station (ISS) – 'The Right Stuff' for astronauts demanded being a team player – an insight immortalized in the phrase 'teamwork makes the dream work' in the opening sentence of the Acknowledgment to the 2017 memoir *Endurance* by Astronaut Scott Kelly (2017), a veteran of the International Space Station who has spent more than 520 days in space.

NASA and its international partners now acknowledge that crew members must not only be technically competent, but also effectively navigate interpersonal interactions in space. Crews have moved beyond the technically competent but socially deficient crews of the Antarctic. A diary entry (Stuster, 2016, p. 78) by a member of the ISS extolled the virtues of the then ISS commander:

> X is a master of good natured fun. I think when he leaves we will see a shift in the enjoyment of the people working the ground jobs. He is brilliant at knowing the perfect balance of fun with professionalism. I am in awe constantly. My love of joking around is immense but I am a mere child next to the talents of my commander. He is gifted.

But space travel is on the cusp of getting even more challenging. We are progressing from long-*duration* space exploration on the ISS (250 miles from earth), to long-*distance* space exploration (LDSE) returning to the moon (250,000 miles away) and then on to Mars (250 million miles away). The acronym LDSE has been used at various times to describe long-duration space exploration, long-distance space exploration, and by the Chinese National Space Administration as Lunar and Deep Space Exploration. We use LDSE here to refer to long-distance space exploration, since the challenges they present – and we seek to model – are beyond just a long-duration mission. It requires the crew to work with much greater autonomy. The days are numbered when we could quip that astronauts are the ‘eyes and ears’ but mission control on Earth remains the ‘brains’ of any mission. The fact that a radio signal can take up to 22 minutes one-way to travel from the Earth to Mars, significantly diminishes the likelihood of a successful resolution in response to a ‘Houston, we have a problem’ call by a Martian crew member. The first words uttered by Capsule Communicator at Mission Control, Charlie Duke, following Armstrong’s confirmation of the down-to-the-wire Apollo 11 landing of the Eagle on the moon was to tell the crew ‘You got a bunch of guys [at mission control] about to turn blue.’ Mission control will not have the luxury to ‘turn blue’ during a Mars landing. The crew will have to coordinate seamlessly on the complex task of landing with unparalleled levels of autonomy from mission control. Future LDSE missions will challenge the frontiers of human collaboration. Crews (representing diverse nations and cultures) are expected to live and work in isolated and confined spaces for up to 30 months, requiring a level of interpersonal compatibility that keeps conflicts between team members manageable and allows team members to rely on one another for support.

While we ponder the substantial unknowns on how to compose dream teams for LDSE, we must leverage what we already know from prior research. We know from prevailing team effectiveness models that teams are best positioned for success when certain enabling conditions are in place (Hackman, 1987, 2012; Mathieu, Maynard, Rapp, & Gilson, 2008; Wageman, Hackman, & Lehman, 2005). Research on team composition, the configuration of attributes among team members, allows us to study the effects of who is selected for a space exploration crew on the future experiences and outcomes of that crew. Team composition models will consider the impact of crew member attributes (e.g. personality, relationships, demographics), but in this context are not just about selecting people for a crew and then washing your hands of the model. Team composition can also consider fluctuation in crew dynamics as they change through the mission as a consequence of crew member attributes.

Team composition is a key enabling structure for teamwork (Bell, 2007). In fact, the composition of the space crew will perhaps be the largest leverage point for mitigating team risk. A vast body of research supports the importance of team composition (Bell, 2007; Mathieu, Tannenbaum, Donsbach, & Alliger, 2014). Team composition is empirically linked to outcomes such as cooperation (Eby & Dobbins, 1997), social integration (Harrison, Price, Gavin, & Florey, 2002), shared cognition (Fisher, Bell, Dierdorff, & Belohlav, 2012), information sharing (Randall, Resick, & DeChurch, 2011), adaptability (LePine, 2005), and team performance (e.g., Bell, 2007).

While it is widely acknowledged that team composition is a critical design feature for effective teams, much of what is known about effective team composition is from research within the confines of conventional workplaces (e.g., production plants). Less is known about how composition affects teams that operate in extreme environments such as those experienced by crews of future space exploration missions. But here we can draw upon insights gathered from teams that share some of the isolated, confined, and extreme (ICE) environments that will confront LDSE. Although they are not exactly comparable, we have learned from contexts such as polar stations, offshore drilling rigs, weather stations, nuclear submarines, and remote construction sites.

While these field and case studies offer important general insights, the extreme environment within which LDSE crews will operate requires carefully designed experiments to study the impact of salient task, social, and physical contextual cues (e.g., isolation, confinement, sleep deprivation) on team functioning. Analog environments such as the Human Exploration Research Analog (HERA) at NASA's Johnson Space Center in Houston, TX and the NEK facility at the Institute for Biomedical Problems in Moscow, Russia are designed to serve as isolated, confined, albeit controlled (ICC) – rather than extreme – environments to mimic some of the realities confronting future space exploration. A number of LDSE-analog studies have examined team composition factors in the LDSE-environment (see Bell et al., 2015 for a review). These studies implicate a number of team composition variables such as gender, national, professional and military background, values, personality, and specific abilities as factors tied to the social integration (e.g., subgrouping, isolation), team processes (e.g., conflict), and emergent states (e.g., shared team mental models) that can affect LDSE mission success. However, many of these studies were correlational, descriptive, and based on small team-level sample sizes. Further they only implicitly recognized that the impact of team composition on functioning was mediated by *social network ties* (such as advice, affect, hindrance, leadership) among crew members. Thus, although team composition is likely to play a critical role in crew social integration, processes, and emergent states for future LDSE crews, the critical team composition factors and the particular patterns of emergent network ties and subsequent outcomes associated with different compositions remain elusive.

The purpose of our chapter is to outline a novel application of computational modeling – and more specifically agent-based modeling – to describe, predict, and prescribe the impact of team composition on team functioning. We report on our use of agent-based modeling to facilitate the study and improvement of crews simulating LDSE as part of a NASA-funded project titled Crew Recommender for Effective Work in Space (CREWS). Specifically, in the next section, we begin by discussing what motivates modeling of social systems and agent-based models (ABMs) of space teams. We trace the use of models in the hard sciences and delineate its use in the social sciences. In subsequent sections, we describe the steps in developing an ABM. We begin by specifying the factors (variables) that influence the construction of an ABM to explore the impact of team composition on crew functioning. Next, we describe how we calibrate these models using empirical data. This requires substantial efforts to instrument the contexts in order to capture all the data needed to calibrate these models. We describe a fairly novel approach to use the data to estimate parameters indexing the effect of various factors in the ABM. Having calibrated an ABM model, we next discuss how to validate the efficacy of the model's predictions. Once validated we demonstrate how these models can be utilized. Finally, we envision how the science described here will translate into action via implementation of a dashboard (or, more accurately, a do-board) to assist decision makers at space agencies such as NASA to anticipate functioning of hypothetical crew configurations prior to a mission, as well as predict – and mitigate – crew functioning post-launch.

COMPUTATIONAL MODELING AND SPACE TEAMS

We begin this section with a brief overview of what we mean by models, and how we use them to aid in composing space teams. A model is a formal representation of a system, real or hypothetical. A simple example of a model would be any mathematical function intended to describe reality, such as a formula from physics describing how a projectile dropped with a certain velocity will change its velocity as it approaches the earth. This model was constructed by physicists in order to detail how existing factors (e.g. the gravity of the earth) and existing theories (e.g. Newton's equations of motion) come together to produce some outcome (e.g. the future speed of the projectile). A model is a way, in very precise language, to describe the process through which some **input** (the original speed with which the object was dropped) becomes translated to some **output** (current velocity of the projectile), as a function of some other **parameters** (e.g. acceleration due to gravity and time elapsed).

Cast in this light, the concept of a model is actually quite broad. A model is any sort of precise, reproducible simplification of reality. Methods such as regression, or a hypothesis being tested in a factorial experimental design, are models, albeit simple ones. There are various ways in which a physicist or an engineer may try to leverage a model. First and foremost, in creating a model, a researcher is required to be precise. They are required to derive an exact, mathematical specification of how they believe each of the variables interacts. As a result, their model is a precise encapsulation of their beliefs that can then be tested, or easily shared with others.

Physicists and engineers rarely stop at simply creating a model. Models are meant to be applied in various ways to explore implications for the phenomenon being modeled. There are three main types of analytics carried out with a model: *descriptive analytics, predictive analytics, and prescriptive analytics* (Delen & Demirkan, 2013). **Descriptively**, a model can provide a lens to describe, understand and/or explain what is observed. In the example, scientists examine how projectiles behave according to the model and begin to test the model using experiments. They estimate realistic values for the parameters, such as the effect of gravity, in order that their model will describe what they observe in real-world experiments. In addition, scientists often leverage the model **predictively**, guessing the future speeds of a hypothetical projectile (even if it was dropped at an initial velocity not previously observed empirically). They speculate on interesting scenarios to test experimentally in the future, and later conduct these experiments to validate whether their model was correct, or how they might revise their model accordingly.

Forecasting is often a valuable end goal of predictive analytics. Consider the case of weather forecasting. However, in many instances, prediction while being a necessary step is not sufficient. While in most instances we tend to grudgingly accept a weather forecast, there are instances where we might want to do something to change it. Consider the case of high profile sporting events such as the Winter Olympics where airplanes are sent to 'cloudseed' a noncompliant weather system to trigger an artificially created 'prescribed' snowfall over the ski routes. Clearly the rarity of this event suggests that weather forecasting doesn't routinely lead to prescriptive analytics. However, in many other areas, once scientists are reasonably comfortable with the performance of their model, they begin to leverage it **prescriptively** in order to make decisions or generate recommendations: how should the inputs (timing or initial velocity) of a projectile be changed in order to obtain a desired outcome (final velocity)? All of these uses – learning about the world, predicting the future, and making the best decision – are jointly tied back to one integrated model that researchers develop.

While these approaches have long been leveraged to understand and enable the physical world, there have been repeated calls to apply these to social systems (see, for example, Pentland, 2014). However, two major hurdles need to be overcome along the way. *First*, unlike most physical systems, social science theory has often not been able to unequivocally identify or decompose the key factors that influence the functioning and outcomes of social phenomena. In the social sciences we do not have – nor are we close to having – the equivalent of an equation that says, given the speed with which a projectile is dropped, the time elapsed and the universal gravitational force of earth, one can instantly predict with high precision the speed of the projectile at any future point in time. Further the distinction between inputs and outputs are often muddied within social systems where they may be interrelated and influencing one another. Our beliefs can influence who we choose to interact with – and who we choose to interact with can influence our beliefs. In modeling parlance, it is very unlikely that rich and complex social phenomena can be adequately modeled using equations that have elegant 'closed form analytical' solutions. Hence the example model from physics we discussed entailed only a deterministic, mathematical calculation, which is largely irrelevant to the social science. When we build models about social processes, it is only natural to incorporate stochastic processes and chance occurrences into our models. After all, not all humans think or interact in the same way each time, and not all influences upon a social process can be perfectly captured by a single model. (See Macy & Tsvetkova, 2015 for an elaboration on the importance of randomness in social science models). In many such cases where we don't really 'know' the model, we need to rely on messier simulation techniques where we have to 'grow' the model. That is, use simulations to model what happens in the system one time step at a time to discover the emergent states of the system at subsequent time points. The *second* hurdle to building predictive and prescriptive models for social science phenomena is closely related to – and indeed an extension of – the first. Even if we were to know the factors that shape a social phenomenon, unlike in the hard sciences, we typically do not have solid evidence about the relative importance of each of these factors. In modeling parlance, we do not know the values of the parameters that provide a quantitative metric by which each factor influences a social outcome. The gravitational constant for acceleration is an example of such a parameter well established in the hard sciences.

We argue that overcoming these hurdles is doable and effective when focused on the effects of composition on space teams. It is able to help us answer questions such as what social networks emerge among crew members? How do crew relationships evolve and change over time? How does one anticipate potential problems that the crew is likely to encounter and what strategies can we prescribe to preempt or mitigate against those problem predictions? Given a pool of potential crew members and role constraints that need to be met, how does one evaluate and rank order the merits of top crew configurations on different dimensions of crew functioning or ability to manage conflict when it occurs? Our preliminary efforts at building and validating these agent-based models of teamwork during simulated space missions to answer the aforementioned questions have been promising. This leads us to believe that further advances with these agent-based models are poised to inform NASA's crew composition questions as it prepares for the Artemis mission that will take the first woman and the next man to the moon in the near future, build the Lunar Gateway, and prepare for a mission to Mars.

This section has outlined the merits of employing models to describe, predict, and prescribe social phenomena. Unlike in the hard sciences, we recognized the limitations for us to 'know' closed-form analytic models to characterize rich social phenomena. Instead we argued for an effort to 'grow' computational models that simulate future states by traversing through time one step at a time. We noted that utilizing these models effectively requires us to overcome two major hurdles – identifying key factors (variables) that influence the social phenomena of interest and estimating the magnitudes of those influences (parameters). Past efforts to overcome these hurdles have relied on expert opinions rather than empirical estimation. But these have limits in situations where experts have divergent opinions on the factors and the magnitude of their impacts. In the next section we delve deeper into how ABMs can help describe, predict, and prescribe interventions for LDSE. We also outline the steps to build, calibrate, validate, and make these agent-based models actionable.

MOTIVATING ABMS FOR SPACE TEAM COMPOSITION

To start developing models of large and complex social systems, we first characterize entities within the system as agents. In our case the **agents** are crew members. The model is a set of probabilistic rules (or equations), which specifies how each agent will update their attitudes (about themselves and other agents) and engage in behaviors (actions and interactions with others). These models often result over time in complex emergent patterns that are not easy for the human mind to intuit although they are entirely derived from probabilistic rules specified by humans. **Agent-based modeling** (ABM) is a perspective on modeling that embraces these ideas to tackle complex problems and understand emergent states.

For those studying teams, ABMs offer an opportunity to examine dynamic team processes. Traditionally, team functions have been studied using Input–Process–Output (IPO) models that focus on how simple main effects result in some sort of outcome in teams. However, there have been increasing calls to move to more nuanced models that incorporate the complex interactions of multiple factors, incorporate emergent states that may form in a team, and incorporate temporal changes in team processes (Grand et al., 2016; Ilgen, Hollenbeck, Johnson, & Jundt 2005; McGrath, Arrow, & Berdahl, 2000). ABMs offer a promising way to bridge this gap. Traditional modeling approaches (regression, factorial design, structural equation models) require researchers to make certain assumptions and test hypotheses that follow a certain structural form. In contrast, agent-based models empower researchers to develop structural patterns of potentially mutual and/or nonlinear influences based on their assumptions. It empowers team researchers to build a more flexible model of the world as they see it.

In the context of space, we have an outstanding opportunity to build a **descriptive** understanding about how various factors (attributes of team members, scheduling of tasks, sleep deprivation, communication delay, lifestyle during LDSE) systemically influence the ability of a crew to collaborate with one another and perform effectively. ABMs of team composition provide a mechanism for researchers to integrate multiple existing theories about team composition, calibrate them with empirical data, and explore the implication of these results.

ABMs are especially well-suited for research in areas, such as LDSE analogs, where we are only able to study a limited number of crews but can collect voluminous amount of data about each of these individual crews, their network relations with one another and how they perform over time. These types of data have traditionally been more amenable for a qualitative, case-driven research approach than quantitative work. Inferential methods often assume a sufficiently large and independently distributed sample that is challenging to gather in LDSE analogs. Furthermore, inferential methods only work toward making 'in sample' claims: data from a 45-day analog mission only describes what to expect from the first 45 days of an LDSE analog, with no strong mechanism to speculate about future trends occurring beyond these 45 days. ABMs address these limitations: They provide an opportunity to build models that can be validated based on high-resolution temporal data collected in other LDSE analogs and projected over longer time spans.

Once a model of how different factors influence crew outcomes in LDSE is constructed, calibrated, and validated, it is now ready to be employed **predictively**. For instance, ABMs allow researchers to conduct *in silico* virtual experiments, in which hypothetical inputs (not previously observed in the real-world) are provided to an ABM to predict what outputs the model will produce. A model that is fed data about crew members' characteristics and their upcoming task schedules can predict potential risks (e.g., interpersonal conflict, high workload) that members of the crew may encounter, paving the way for mission support to plan future countermeasures aimed at mitigating these risks.

Finally, ABMs have **prescriptive** uses that can help mission support to plan those future countermeasures aimed at mitigating those aforementioned risks. Prescriptive analytics will evaluate the efficacy of these options. Relatedly, given the state of the crew, ABMs can recommend (or prescribe) how tasks can be scheduled, based on workload, sleep deprivation, or other factors, in a way that will help astronaut crews operate at their optimal performance. As such ABMs will be a potentially valuable tool to help researchers offer operational assistance to shape the effectiveness of team processes in LDSE.

DEVELOPING AN AGENT-BASED MODEL FOR CREW COMPOSITION EFFECTS

The development of an agent-based model is a complicated and iterative process, in which researchers apply many different techniques to create, improve, and learn from their model. We outline steps we used to develop an agent-based model of team composition by describing four key processes we carried out: **model construction**, **model calibration**, **model validation**, and **model application** (Figure 6.1). While we apply this approach to team composition, it can be applied to other dynamic phenomenon in LDSE analog research.

In *model construction*, we specify the system of interdependent variables of interest that capture the social phenomena we want to explain. We relied on theory, prior empirical research, and meta-analyses in order to select variables to include in our model and to specify potential mechanisms by which these variables may influence one another. The *model calibration* stage is where the empirical data collected in analogs are used to estimate the parameters of the model. In the *model validation* stage we evaluate the extent to which the model is valid in terms of fitting the observed data on which it was trained as well as on new test data. Finally in the *model application* stage, we conduct virtual experiments to predict what might happen in a hypothetical team as well as evaluate various prescriptive actions to mitigate potential problems that are predicted. We hasten to add that there is no single ‘correct’ approach to developing an agent-based model. Despite its linear

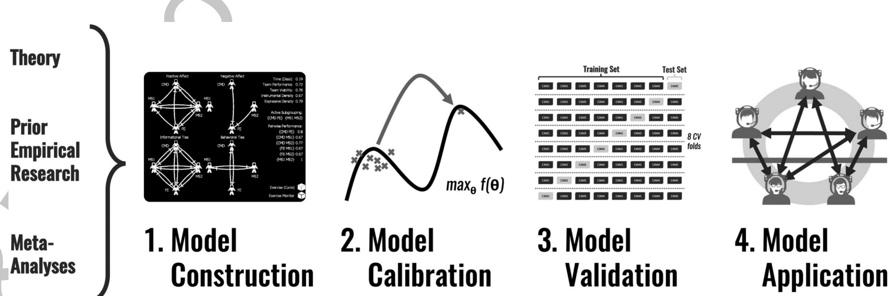

FIGURE 6.1 Flowchart for the steps that may be used in developing an emulative ABM.

representation, in practice, model development is an iterative process of refinement and extension – moving through each process multiple times and adapting plans for the next step based on what happened in the previous ones.

Model Construction

Defining Model Scope

The first step in constructing an agent-based model is to describe the models' scope: Who are the agents, what are the output metrics of the model that we seek to explain (e.g. team functioning, performance, viability) and what factors influence, and are perhaps in turn influenced by, these output metrics? These questions form the foundation of what the model will try to accomplish, and how it will go about doing it.

Until recently, because of the paucity of dynamic empirical data, ABMs were more heavily utilized to develop simple, stylized models of social phenomena and were used primarily to explore how changes in inputs or mechanisms might impact emergent outcomes. For instance, a simple stylized model where new agents entering a network were more likely to connect with already well-connected nodes demonstrated the plausibility of preferential attachment as a theoretical mechanism to explain the widespread prevalence of scale-free 'hub-and-spoke' social networks (Wilensky, 2005). Models designed to puzzle through such thought experiments are often referred to as *intellective* computational models (Mavor & Pew, 1998). The parameters in these computational models are often arbitrarily chosen with little loss of generalizability. However, with the increasing availability of high-resolution temporal data, there is greater interest in the development of *emulative* computational models (Carley & Hirshman, 2011). These much larger models seek to emulate in substantial detail the dynamic features and empirical characteristics of a specific team or organization (Carley, 2009). They often have, by comparison, a much larger number of inputs and outputs; however, the availability of large amounts of dynamic empirical data eliminate the need for modelers to *a priori* specify parameters for the impact of these variables on the phenomena of concern. Instead we use novel genetic algorithms and optimization techniques to empirically estimate these parameters (Stonedahl & Wilensky, 2010a; Sullivan, Lungeanu, DeChurch, & Contractor, 2015; Thiele, Kurth, & Grimm, 2014). Using empirical data to estimate the parameters in a computational model is a novel contribution to ABM research. The idea is somewhat analogous to a statistical (e.g., regression) model, in which empirical data is employed to identify whether, and to what extent, variables influence one another. Using empirical data to estimate parameters in ABM have the potential to blunt criticism that modelers face from theorists or empiricists who are wary of believing insights drawn from computational models which include, arguably, arbitrarily specified parameters – rather than parameters supported by empirical data.

Having decided on the agents, the decision to design (in our case) an emulative (rather than intellective) model, and high-level categories of inputs and outputs, the next step is to develop the agent-based model.

Theory, Prior Empirical Research, and Meta-Analysis

In the first step, we used theory, prior empirical research, and meta-analyses for two purposes: (i) to identify a system of variables that are interrelated with the phenomena of interest, and (ii) create probabilistic rules that specify how agents' attitudes and behaviors shape, and are shaped by, the system of variables. In our case the outcomes of interest are crew performance and viability. However, a central, arguably idiosyncratic, premise of our modeling effort is that the impact of compositional factors on crew performance and viability is completely mediated by crew members' network relations (Figure 6.2).

Indeed, a wide-body of extant literature (Balkundi & Harrison, 2006; Crawford & LePine, 2013; Mehra et al., 2006) have established significant connections between social relations and measures of team performance We identified four social relations that were relevant to analog research – task affect, task hindrance, leadership, and followership. In addition, our research on HERA crews has shown that properties of the task affect, task hindrance, leadership, and followership networks were all correlated with objective measures of performance on team tasks (Antone et al., 2019).

Given our premise that social relationships mediate the effects of team composition on crew performance and viability, the remainder of the model is focused on compositional, network, and environmental factors that influence social relationships among crew members. Figure 6.2 provides a schematic of the factors in our ABM influencing social relationships among crew members. This model was based on a review of the theoretical and empirical literature on team composition, a smaller subset of case studies that looked at teams in isolated and confined environments and meta-analyses on team composition.

An Integrated Model of Team Composition

The factors shaping social relationships among crew members fall into five buckets: First, we consider the endogenous effects labelled 'Social Network Trends' in Figure 6.3. These include temporal patterns such as inertia – the likelihood of a crew member enjoying working with another in the future is often best predicted by the extent

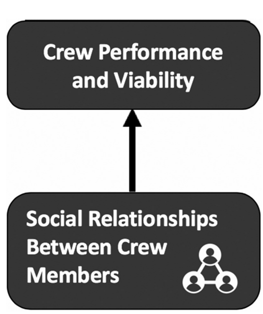

FIGURE 6.2 Core networks predicting key outcomes of performance and viability.

to which the crew members currently enjoy working with one another. Another common endogenous mechanism is based on reciprocity. If a crew member enjoys working with another, it is likely that the other will also report enjoying working with the former. Likewise, crew relation may be transitive, if A looks to B for leadership and B looks to C for leadership, A might also look to C for leadership. Finally, crew relations might exhibit the emergence of hubs. One crew member might draw hindrance ties from all other members.

The two buckets on the right consider the compositional effects of individuals' personality on crew relations. The bucket labelled 'Personality' considers the extent to which a crew member's personality characteristics (Five Factor Model personality traits and facets, values, coping styles, psychological collectivism, and self-monitoring) make them more (or less) likely to report (or receive) specific social ties from other crew members. The bucket labelled 'Personality Fit' considers the extent to which the match (or mismatch) in personality characteristics between two crew members might increase or decrease the likelihood of a social relation between them.

The two buckets on the left side of Figure 6.3 consider environmental factors that influence crew social relations. The bucket labelled 'ICC' refers to the impact of contextual factors (Isolation, Confinement, and Controlled conditions) on crew social relationships. Finally, the bucket on the bottom left labeled 'Tasks and Scheduling' considers how aspects of the tasks impact crew relations. Specifically, we modeled the extent to which crew relations were influenced by the workload, interdependence, situational strength, and duration of each task the crew carried out.

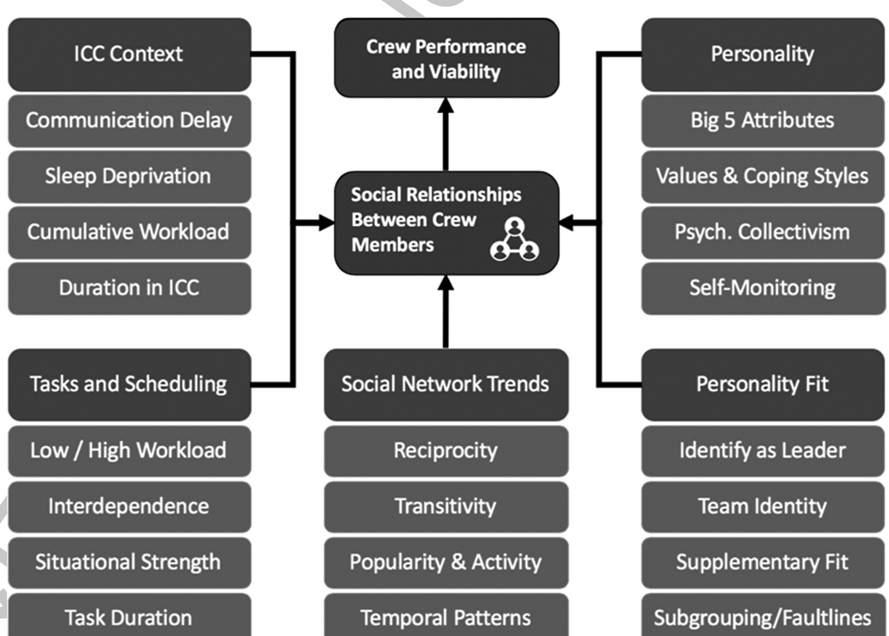

FIGURE 6.3 Factors integrated into our ABM of Teamwork in LDSE.

Each of these potential influences were codified as a system of probabilistically driven rules that would update crew relations at each time point based on prior time points for the entire duration of the 30- or 45-day missions. Simplifying assumptions are made about the level of change during sleep periods. Time invariant factors such as personality and personality fit would have a baseline effect across all time periods while time variant factors such as days in isolation, variations in tasks and scheduling, and fluctuations in the social relations themselves had a more dynamic impact on future states of social relations. These systems of equations were then implemented in Netlogo (Wilensky, 1999), a widely used ABM platform. The ABM model was now ready to be calibrated as described in the following section.

Model Calibration

A distinctive feature of our deployment of ABM is to rely entirely on empirical data to estimate the magnitude with which each factor in our agent-based model influenced crew relations. This is in stark contrast with most prior ABM efforts (see Sullivan et al., 2015 for an exception) where the researcher uses some heuristic (a literature review of effect sizes or expert opinion) to specify the magnitude with which various factors impacted outcomes. As Smith and Rand (2017) argued, using data generated from real experiments is the ideal method to design and calibrate agent-based model's rules and the mechanisms.

Collecting high-resolution data for the study of long-duration space exploration is a major challenge. While it is not possible to intensely survey and monitor actual crews in space, we relied on data gathered in NASA's Human Exploration Research Analog (HERA) at Johnson Space Center. HERA simulates long-duration space missions with a crew of four ranging for a period of 30–45 days. HERA places crews of individuals in conditions that simulate space exploration: completing simulated tasks, living in a small module for extended periods, experiencing communication delays with mission control as they 'travel' away from earth, as well as designated periods of extended sleep deprivation. Data collected in isolated and confined environments such as HERA is arguably the closest alternative for studying crews to actual space missions.

That said, these long-duration space exploration analogs are also expensive and time-consuming to operate. Researchers are only afforded the opportunity to observe a handful of missions every year. The upside is that for the crews that are observed, we can observe many variables over time. For our model calibration, we obtained data from eight separate four-person crews completing 30–45 day missions in the HERA analog operated by NASA.

To have our model estimate parameters based on what occurs in these HERA crews, we must collect data on all variables identified in Figure 6.2. Time invariant personality and personality fit variables only needed to be collected once using standard psychometric scales. Time variant variables needed to be measured at several points in time. The latter included social networks elicited from the crew via sociometric surveys at eight points in time over the course of a 30-day mission or 12 points in time over the course of a 45-day mission. In addition, we were able to collect data using pre-mission and post-mission surveys.

As dependent variables in our model, we included measures of four relational networks: task affect, task hindrance, leadership, and followership. These four networks capture a long-standing distinction in the small group literature on task and social needs. The task affect and hindrance capture positive and negative working relationships among crew members. Task affect was measured with the prompt: 'With whom do you enjoy working?' Task hindrance was elicited with the prompt: 'Who makes tasks difficult to complete?' In addition to assessing manifest social relations, we also included two networks capturing behavioral and motivational aspects of teams: leadership and followership. Leadership was elicited by asking 'To whom do you provide leadership?' Followership relations were assessed by asking: 'Who do you rely on for leadership?' These four prompts yield four directed networks, each examined in relation to performance. We also coded task characteristics based on crew members' perceptions of workload and we were also provided detailed minute-by-minute task schedules (nicknamed the 'playbook') for individuals working by themselves or in teams over the course of the entire mission. Finally, we were able to design our own tasks carried out by the HERA crews to gauge multiple measures of team performance across task types (Larson et al., 2019; Antone et al., 2020).

To estimate the parameters of the ABM model, we used genetic search algorithms implemented in the BehaviorSearch tool for NetLogo (Stonedahl & Wilsensky, 2010c). The BehaviorSearch tool allows for the specification of an objective function that is minimized or maximized according to some set of constraints to 'calibrate' the model. Calibration simply describes the process of manipulating a model to get closer to a desired behavior (Calvez, & Hutzler, 2005; Stonedahl & Wilensky, 2010b). In this case, the desired behavior is matching as closely as possible the simulated social relations among crew members with the empirical observed social relations among crew members. The objective function we chose was the mean squared error between simulated crew relations and empirical crew relations. The BehaviorSearch software implements several search algorithms, which can be used to find a set of parameters that minimizes the mean SSE. To find the parameters for this model, each of the different search algorithms were tested. In our case, the standard genetic algorithm yielded the best results. Our results indicated, for instance, that crew members tend to enjoy working with individuals who are high on self-monitoring. Further, these individuals are less likely to be viewed as making tasks difficult to complete. Further, high workload schedules make crew members less likely to enjoy working with others. Turning to leadership relationships, our model estimates indicate that two crew members are not likely to claim leadership over one another. However, when crew members rely on one another for leadership, it is likely to be reciprocated.

Unlike traditional statistical inferential techniques, estimates obtained from BehaviorSearch algorithms are not accompanied with standard errors and hence are not amenable to standard significance tests. However, to assess the robustness of the parameters estimated for, say, parameter P, we run the model fixing all the other parameters to the values estimated by BehaviorSearch, while letting the parameter P vary over its range (from –1 to 1) using enough replications to compute the mean fit error. For example, to test the significance of the finding that crewmembers tend to enjoy working with individuals who are high on self-monitoring, we ran the model

500 times using the parameters determined by BehaviorSearch (e.g., self-monitoring parameter for the recipient of task enjoyment relations was 0.56). Then, we ran the model 500 times with all the same parameters except the self-monitoring parameter that could vary over its range (from –1 to 1). Finally, a one sample *t*-test was performed to determine whether the set of errors estimated with the fit parameter (0.56) are less than those estimated by allowing the focal parameter to vary (from –1 to 1). A negative and significant effect means that the focal (in this case, self-monitoring) parameter has a measurable and significant effect on reducing the error for crew social relations; as such it plays a significant role in matching the social relations in the simulated and empirical model. The procedure is repeated for all parameters estimated.

Model Validation

Having a calibrated model with parameter estimates begs the inevitable next question. How well did we do? The next phase is validation, in which our goal is to assess the extent to which simulation results from our agent-based model provides a useful reflection of observed data. There are three types of validation on which we focus: We confirm **face validity**, the extent to which the variables and mechanisms make intuitive sense for the phenomenon we are modeling, by relying on extant theory. Because we are producing a model to mimic reality, our goal is to check that the structure of our model is reasonable, before moving onto empirical approaches for assessing validity. For instance, we would expect that at least some of the parameter estimates for variables impacting crew relations have theoretical plausibility. Consider the result we reported in the previous section that workload schedules make crew members less likely to enjoy working with others. While not groundbreaking, results such as these help confirm the face validity of the model and open up the possibility for taking seriously, and puzzling over, some potentially counter-intuitive estimates.

We next seek to confirm **internal validity**, the extent to which our model can explain what happens in the data we empirically observed. Specifically, we conduct direct comparisons between our predicted and simulated results for the same data set. Alongside face validity, these tests determine the extent to which the rules in the model are able to generate patterns in the simulated data that are aligned with the observed data. For instance, we examine plots of the number of relations for each crew in our simulations, in comparison with their observed values, as well as the predictive performance of our model at different points in time. Overall, we confirm that our model tends to mirror the aggregate trends in the data used to estimate it.

Finally, we consider issues of **external validity**. A key question, for an emulative agent-based model in particular, is how well the model performs at making predictions for an unobserved crew? With a limited sample of crews, the best approach to estimating the predictive performance of our model is through cross validation. Given we have observed eight independent crews, we perform eight-fold cross validation: We select one crew to hold out as a test set, estimate our models' parameters using data from the remaining seven crews, and then use this set of parameters to simulate the held-out crew. These simulated ties are compared with the empirically observed data to evaluate predictive performance. By repeating this process eight times, using each crew as the test set once, we obtain an estimate of how well our model would predict relations for a future crew.

To evaluate our model, we examine the confusion matrix cross tab between presence or absence of predicted and observed ties, alongside summary statistics such as accuracy, precision, recall, F_1 scores, ROC curves, and precision–recall curves (Davis & Goadrich, 2006; Fawcett, 2006). These summary measures provide a better understanding of model quality than accuracy alone, especially in the case where the relationship being predicted is either very frequently present, or very frequently absent. For instance, in our data, task affect relations are present 81.3% of the time, and task hindrance ties occur only 23.3% of the time. In this case, a trivial classifier predicting that all task affect ties exist and no task hindrance ties exist would obtain deceptively impressive but fundamentally useless accuracy scores of 81.3% and 76.7%, respectively. Such a classifier would not be useful practically in distinguishing who is likely to have a certain tie. Therefore other performance metrics, beyond accuracy, must be assessed.

Specifically, we compute (1) Precision scores which indicate the percent of predicted ties that were observed in real crews, (2) Recall scores which indicate the percent of observed ties that were correctly predicted by our model, and (3) F_1 scores, which use the harmonic mean of precision and recall as a measure of performance. Results of our model validation for predicting 'who crew members enjoy working with' achieved average F_1 scores of 0.85 for internal validity (on the training data set) and average F_1 scores of 0.81 for external validity (on a test data set). However, the results of our model validation for predicting who crew members cite as 'making tasks difficult to complete' (i.e. hindrance ties), our average F_1 scores for internal validation fell to 0.56 and for external validation fell to 0.37. The disparity in validity between the two types of social relations is, at least in part, an artifact of the relatively sparse number of observed hindrance ties as compared with task affect ties, thus making it more difficult to capture that signal adequately.

With small-sample data, cross-validation testing is critical to ensure we are not overfitting our model to nongeneralizable specifics of our observed crews. Additionally, such estimates of performance are necessary when assessing whether our model will be able to make predictions of sufficient quality to be used in practice. This type of validation, in particular, identification of uncertainty in predictions, has been considered critical by NASA in its published Standards for Models and Simulations (Steele, 2007; NASA Standard, 2009).

The greatest challenge we will encounter, in modeling space exploration, however, is our reliance on analog data. What we observe in 30- to 45-day analog missions will not fully reflect the empirical realities of LDSE, and thus our findings may not completely generalize to these crews. Cross-validation testing cannot account for these issues. As we work toward building models usable for real-world decision-making, there is a need to start testing analog models outside of HERA – testing our models in scenarios involving longer missions, more extreme environments, different types of work, and multinational crews. Assessing generalizability in a varied ensemble of LDSE analogs (e.g. Antarctic studies, SIRIUS and HI-SEAS analogs) will be the best we can do prior to working on actual space missions.

Model Application

Virtual Experiments

We began the process of constructing an ABM with the selection of variables informed by prior theory and research. The ABM we constructed was then calibrated using empirical data collected specifically to test this model. Next, we validated the ABM to assess how well the data simulated from our ABM aligned with the data used to calibrate it and subsequently how well it predicted crew relations for an out-of-sample data set that was not used to calibrate it. Once the ABM passes muster through these three stages (construction, calibration, and validation), it is ready to be deployed for the final stage of *model application*. As mentioned earlier, we conduct virtual experiments at the *model application s*tage to predict what might happen in a hypothetical team as well as evaluate various prescriptive actions to mitigate potential problems that are predicted.

Starting with HERA Campaign 5 in early 2019, we have been conducting virtual experiments to predict in-mission crew dynamics in HERA missions based only on pre-mission data we collect about the composition of the crew. These virtual experiments allow us to predict *in silico* the dynamics for a crew that has not actually deployed but is based on an ABM calibrated and validated with other crews. We use the results of these virtual experiments to identify which crew relations might reach dysfunctional levels and when during the mission this is likely to occur. Once these potential pain points have been predicted, we use virtual experiments to explore *prescriptive* strategies to mitigate against them. One arrow in our quiver of mitigation strategies is the task schedule. As mentioned previously, HERA crews like their counterparts in space, work on a strictly regimented task schedule. The 'playbook' assigns specific time slots each day for the completion of solo as well as tasks assigned to pairs, three members or the entire crew. In the event of a potential relational issue between two crew members, we run virtual experiments where we keep everything the same except making tweaks to the schedule of which crew members are paired with one another and on which tasks. For instance, we might run a virtual experiment to see if a good mitigation strategy might be to not schedule tasks for a specific crew pairing as part of a 'cooling-off' period. Alternatively we can conduct virtual experiments that schedule tasks for these two crew members with a third member they both enjoy working with. Yet another mitigation strategy we explore is to pair them only on tasks at which they excel to explore if joint success on the task repairs the relationship.

TRANSLATING SCIENCE TO PRACTICE

So far, we have described the steps by which agent-based models are developed and evaluated for their predictive and prescriptive capabilities. However, our ultimate goal is to produce models that are able to be used by actual decision makers. In anticipation of that eventuality, we have developed a prototype dashboard for use pre-mission by decision makers for crew selection and in-mission by decision makers for planning and operations. The dashboard called TEAMSTAR (**T**ool for **E**valuating **A**nd **M**itigating **S**pace **Tea**m **R**isk) has one fundamental goal: make insights from our ABM accessible to decision makers without them requiring any knowledge of agent-based modeling. As such TEAMSTAR aspires to be both a dashboard – and a 'do-board.'

TEAMSTAR is powered at the back-end by ABM and requires the administrator to upload relevant data (e.g., attributes of potential team members, prior relations, task schedules).

Prior to the mission, TEAMSTAR provides decision makers with an easy to use interface to predict how a hypothetical team's social relations are likely to evolve over the course of a mission. The decision maker selects a pool of potential crew members and then composes hypothetical teams by simply binning names of hypothetical teams (Figure 6.4). TEAMSTAR runs the virtual experiments in the background and provides decision makers with predictions about the relationships between crew members at any point in time over the upcoming mission (Figure 6.5).

To be useful, a predictive team composition model needs to be flexible in terms of staffing capabilities, and its ability to estimate risks associated with different hypothetical crews. First, different staffing strategies can be used when composing teams. One strategy is for the compatibility of all crewmembers to be considered simultaneously. Another strategy is to first identify critical team members (e.g., the commander) and then assess the remaining crew members' compatibility with those critical members. Because LDSE-crews are expected to be multinational, there may be little ability to influence the decision to select all team members, and instead the compatibility of a particular individual or set of individuals will need to be considered. Thus, a predictive team composition model needs to be flexible in its ability to inform different staffing strategies.

The ABM powering TEAMSTAR will enable decision makers to evaluate composition scenarios for an entire set of teams, for single-member replacements, and/or for subsets of teams. This will maximize its utility given that, in international missions, only some of the astronauts will be selected by NASA. TEAMSTAR

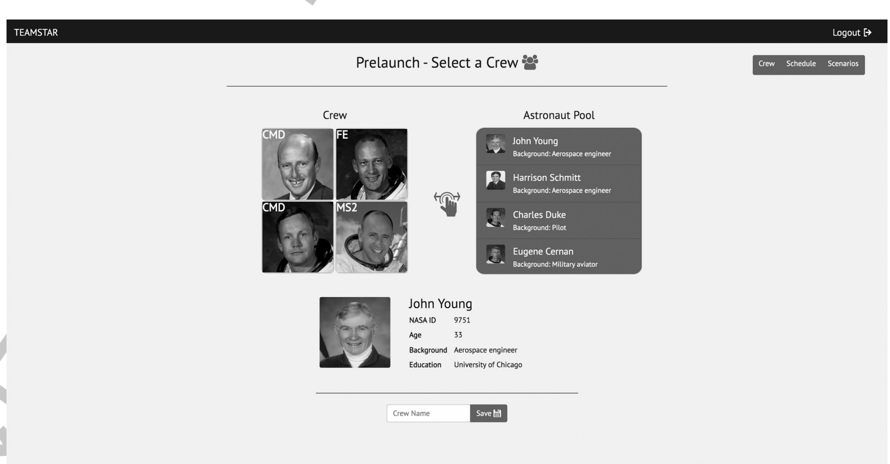

FIGURE 6.4 Selecting hypothetical crews to predict their dynamics.

FIGURE 6.5 Predicting team dynamics for a hypothetical team pre-launch.

can also be useful in re-staffing teams should a member be replaced during pre-mission training, recommending a best replacement member to NASA from a set of alternatives.

Once in-mission, TEAMSTAR projects how the team is likely to evolve in terms of risk markers such as social integration, team processes (e.g., conflict), and emergent states (e.g., shared mental models). Since the ABM is both temporal and relational in nature, TEAMSTAR also produces detailed results on what social relations and overall crew cohesion looked like in the past and will look like in the future, with confidence intervals for these predictions (Figure 6.6). Second, because there may be constraints on the ability to influence the team's composition as a whole,

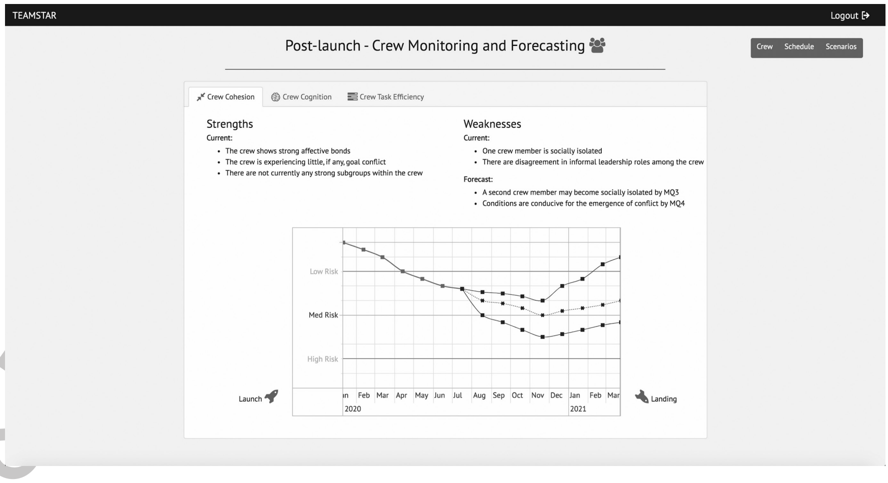

FIGURE 6.6 Past and projected trends of a crew pathway into the mission.

it is important to understand the risks associated with the team's composition. With a predictive model of team composition, different risks (e.g., subgrouping, conflict, difficulty maintaining shared mental models) can be estimated for proposed or current crew compositions. Personalized medicine acknowledges that not all humans have the same needs; these individualized needs should provide the basis for countermeasures in human space flight (Schmidt & Goodwin, 2013). In the same way, not all crews will have the same needs. Estimated risks from the predictive model of team composition can be used to understand the training needs of a specific crew and guide the development and strategic application of countermeasures. In-flight countermeasures could be mapped to specific crew compositions and risks. For example, for a crew composition that has a high risk for subgroup conflict across national background, mission control could provide 'critical' work, specifically encouraging members from different subgroups to work interdependently, at key points in the crew's life cycle.

CONCLUSION

This chapter has sought to introduce how an agent-based modeling approach can be used to describe, predict, and prescribe the consequences of team composition: We have described the development of an emulative agent-based model of social relations in crews, illustrating the process of model construction, model calibration, model validation, and model application.

We recommend the following resources for those interested in learning more about agent-based modeling processes (Wilensky & Rand, 2015; Gilbert, 2007; Heath, Hill, & Ciarallo, 2009), software for implementing agent-based models (e.g. Netlogo, Repast), and approaches for estimating and validating agent-based models (Thiele, Kurth, & Grimm, 2014). A future direction, for models such as ours, may be better quantification of the statistical uncertainty around model parameters. In particular, Bayesian approaches have been identified as promising for extremes team research, due to their ability to represent uncertainty and incorporate extant prior knowledge into these assessments (Bell et al., 2018).

Our model is not without limitation. In developing a model for space exploration, we struggled with choices between constructing models that were more exhaustive, or more selective, in their scope. There is, naturally, a desire for researchers to build more integrative models. If more variables and mechanisms are included in a model, more nuances can be represented, and the influences of all these variables and mechanisms can fully be considered when using the models for prediction or decision-making. However, in the presence of a finite sample of data, including too many related or correlated variables can diminish our certainty about the 'true' or 'best' value of the model parameters for each one. This trade-off will be a key consideration for all models developed for space exploration teams. As an oft quoted statistical aphorism states, 'all models are wrong but some are useful' (Box, 1979). We will never have a perfect model for space crew composition, but hopefully we can keep building better models that are highly useful.

Overall, we have demonstrated a proof-of-concept of the potential role that agent-based models could serve in helping prepare future crews for long-duration space exploration. We hope that this work lays the foundation for future researchers or practitioners interested in developing agent-based models for space exploration crews. As more and more data is gathered from space exploration analogs, progressively more nuanced agent-based models can be developed for space exploration.

One final note: Over the past six decades, research conducted for space missions have had significant knowledge spillover in various sectors back on Earth. For instance, we have NASA to thank for the cordless drills originally designed to help astronauts drill on the surface of the moon. High-intensity LED (light emitting diodes) were developed for the NASA shuttles, but are now making great advances in power efficiency back on Earth. Astronauts needed something to keep their recycled water clean. NASA invented a filter with activated charcoal to neutralize pathogens. These technologies are used extensively around the world, including the Global South. Remarkably, all of these innovations have spun out of technological and health challenges faced in space. Today we are on the brink of an innovation that will have spun out of a social science challenge – anticipating and mitigating social dynamics in teams. Alongside important conversations about ethics and privacy, we are beginning to see interest in deploying advanced people analytics, especially relational analytics (Leonardi & Contractor, 2018) that will extend the models and methodology developed for space missions and apply them to the changing nature of work here on earth – and perhaps some day in interplanetary work contexts.

ACKNOWLEDGMENTS

The material is based upon work supported by NASA under award No. NNX15AM32G. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

REFERENCES

Antone, B., Gupta, A., Bell, S., DeChurch, L., & Contractor, N. (2019). Testing influence of network structure on team performance using STERGM-based controls. In H. Cherifi, S. Gaito, J. Mendes, E. Moro, & Rocha L. (Eds.), *Complex networks and their applications VIII. Complex networks 2019. Studies in computational intelligence* (Vol. 882). Cham: Springer.

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal*, *49*(1), 49–68.

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of applied psychology*, *92*(3), 595.

Bell, S. T., Brown, S. G., Abben, D. R., & Outland, N. B. (2015). Team composition issues for future space exploration: a review and directions for future research. *Aerospace Medicine and Human Performance*, *86*(6), 548–556.

Bell, S. T., Fisher, D. M., Brown, S. G., & Mann, K. E. (2018). An approach for conducting actionable research with extreme teams. *Journal of Management*, *44*(7), 2740–2765.

Biersner, R. J., & Hogan, R. (1984). Personality correlates of adjustment in isolated work groups. *Journal of Research in Personality*, *18*(4), 491–496.

Box, G. E. (1979). Robustness in the strategy of scientific model building. In R. Launer & G. Wilkinson (Eds.) *Robustness in statistics* (pp. 201–236). New York, NY: Academic Press.

Calvez, B., & Hutzler, G. (2005, July). Automatic tuning of agent-based models using genetic algorithms. In *International workshop on multi-agent systems and agent-based simulation* (pp. 41–57). Berlin, Heidelberg: Springer.

Carley, K. M. (2009). Computational modeling for reasoning about the social behavior of humans. *Computational and Mathematical Organization Theory*, *15*(1), 47–59.

Carley, K. M., & Hirshman, B. (2011). Agent based model. In G. A. Barnett (Ed.), *Encyclopedia of social networks* (Vol. I, pp. 11–16). Thousand Oaks, CA: SAGE Publications.

Crawford, E. R., & LePine, J. A. (2013). A configural theory of team processes: Accounting for the structure of taskwork and teamwork. *Academy of Management Review*, *38*(1), 32–48.

Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In Proceedings of the International Conference on Machine Learning (pp. 233–240). New York, NY: ACM.

Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems*, *55*(1), 359–363.

Eby, L. T., & Dobbins, G. H. (1997). Collectivistic orientation in teams: An individual and group-level analysis. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *18*(3), 275–295.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874.

Fisher, D. M., Bell, S. T., Dierdorff, E. C., & Belohlav, J. A. (2012). Facet personality and surface-level diversity as team mental model antecedents: implications for implicit coordination. *Journal of Applied Psychology*, *97*(4), 825.

Gilbert, N. (2007). *Agent-based models* (Vol. 153). Thousand Oaks, CA: SAGE Publications.

Grand, J. A., Braun, M. T., Kuljanin, G., Kozlowski, S. W., & Chao, G. T. (2016). The dynamics of team cognition: A process-oriented theory of knowledge emergence in teams. *Journal of Applied Psychology*, *101*(10), 1353–1385

Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice-Hall.

Hackman, J. R. (2012). From causes to conditions in group research. *Journal of Organizational Behavior*, *33*(3), 428–444.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface-and deep-level diversity on group functioning. *Academy of Management Journal*, *45*(5), 1029–1045.

Heath, B., Hill, R., & Ciarallo, F. (2009). A survey of agent-based modeling practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation*, *12*(4), 9.

Huntford, R. (2013). *Shackleton*. London: Hachette.

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.

Kelly, S. (2017). *Endurance: My year in space, a lifetime of discovery*. New York, NY: Knopf.

Larson, L., Wojcik, H., Gokhman, I., DeChurch, L., Bell, S., & Contractor, N. (2019). Team performance in space crews: Houston, we have a teamwork problem. *Acta Astronautica*, *161*, 108–114.

Leonardi, P., & Contractor, N. S. (2018, November–December). Better people analytics: Measure who they know, not just who they are. *Harvard Business Review*, *96*, 70–81.

LePine, J. A. (2005). Adaptation of teams in response to unforeseen change: Effects of goal difficulty and team composition in terms of cognitive ability and goal orientation. *Journal of Applied Psychology*, *90*(6), 1153.

Macy, M., & Tsvetkova, M. (2015). The signal importance of noise. *Sociological Methods & Research*, *44*(2), 306–328.

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997–2007: A review of recent advancements and a glimpse into the future. *Journal of management*, *34*(3), 410–476.

Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and integration of team composition models: Moving toward a dynamic and temporal framework. *Journal of Management*, *40*(1), 130–160.

Mavor, A. S., & Pew, R. W. (1998). *Modeling human and organizational behavior: Application to military simulations*. Washington, DC: National Academy Press.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychology Review*, *4*(1), 95–105.

Mehra, A., Dixon, A. L., Brass, D. J., & Robertson, B. (2006). The social network ties of group leaders: Implications for group performance and leader reputation. *Organization science*, *17*(1), 64–79.

NASA (2009). NASA standard for models and simulations. NASA STD-7009. Hampton, VA: Langley Research Center.

Natani, K., & Shurley, J. T. (1974). Sociopsychological aspects of a winter vigil at South Pole station. *Human Adaptability to Antarctic Conditions*, *22*, 89–114.

Pentland, A. (2014). *Social physics: How good ideas spread-the lessons from a new science*. New York, NY: Penguin.

Randall, K. R., Resick, C. J., & DeChurch, L. A. (2011). Building team adaptive capacity: The roles of sensegiving and team composition. *Journal of Applied Psychology*, *96*(3), 525.

Schmidt, M. A., & Goodwin, T. J. (2013). Personalized medicine in human space flight: Using Omics based analyses to develop individualized countermeasures that enhance astronaut safety and performance. *Metabolomics*, *9*(6), 1134–1156.

Smith, E. B., & Rand, W. (2017). Simulating macro-level effects from micro-level observations. *Management Science*, *64*(11), 5405–5421.

Steele, M. J. (2007, July). The NASA standard for models and simulations. In *Proceedings of the Computer Simulation Conference* (pp. 5–10). San Diego, CA: Society for Computer Simulation International.

Stonedahl, F., & Wilensky, U. (2010a, May). Finding forms of flocking: Evolutionary search in abm parameter-spaces. In *International workshop on multi-agent systems and agent-based simulation* (pp. 61–75). Berlin, Heidelberg: Springer.

Stonedahl, F., & Wilensky, U. (2010b, November). Evolutionary robustness checking in the artificial Anasazi model. In *2010 AAAI Fall Symposium Series*. Arlington, VA: AAAI.

Stonedahl, F., & Wilensky, U. (2010c). BehaviorSearch [computer software]. *Center for Connected Learning and Computer Based Modeling, Northwestern University, Evanston, IL*. http://www.behaviorsearch.org.

Stuster, J. (2016). *Behavioral issues associated with long-duration space expeditions: Review and analysis of astronaut journals experiment Phase 2* (NASA/TM-2016-218603). Houston, TX: NASA Johnson Space Center.

Sullivan, S., Lungeanu, A., Dechurch, L., & Contractor, N. (2015). Space, time, the development of shared leadership networks in multiteam systems. *Network Science*, *3*(1), 124–155.

Thiele, J. C., Kurth, W., & Grimm, V. (2014). Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Social Simulation*, *17*(3), 11.

Wageman, R., Hackman, J. R., & Lehman, E. (2005). Team diagnostic survey: Development of an instrument. *The Journal of Applied Behavioral Science*, *41*(4), 373–398.

Wilensky, U. (1999). *NetLogo*. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Wilensky, U. (2005). NetLogo preferential attachment model. *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL*. http://ccl.northwestern.edu/netlogo/models/PreferentialAttachment.

Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. Cambridge, MA: MIT Press.

Wolfe, T. (1979). *The right stuff*. New York: Farrar, Straus and Giroux.