



# How Personality and Communication Patterns Affect Online *ad-hoc* Teams Under Pressure

Federica Lucia Vinella<sup>1\*</sup>, Chinasa Odo<sup>2</sup>, Ioanna Lykourantzou<sup>1</sup> and Judith Masthoff<sup>1</sup>

<sup>1</sup> Human Centred-Computing, Information and Computing Sciences, Utrecht University, Utrecht, Netherlands, <sup>2</sup> The School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, United Kingdom

Critical, time-bounded, and high-stress tasks, like incident response, have often been solved by teams that are cohesive, adaptable, and prepared. Although a fair share of the literature has explored the effect of personality on various other types of teams and tasks, little is known about how it contributes to teamwork when teams of strangers have to cooperate *ad-hoc*, fast, and efficiently. This study explores the dynamics between 120 crowd participants paired into 60 virtual dyads and their collaboration outcome during the execution of a high-pressure, time-bound task. Results show that the personality trait of Openness to experience may impact team performance with teams with higher minimum levels of Openness more likely to defuse the bomb on time. An analysis of communication patterns suggests that winners made more use of action and response statements. The team role was linked to the individual's preference of certain communication patterns and related to their perception of the collaboration quality. Highly agreeable individuals seemed to cope better with losing, and individuals in teams heterogeneous in Conscientiousness seemed to feel better about collaboration quality. Our results also suggest there may be some impact of gender on performance. As this study was exploratory in nature, follow-on studies are needed to confirm these results. We discuss how these findings can help the development of AI systems to aid the formation and support of crowdsourced remote emergency teams.

**Keywords:** crowdsourcing, collaboration, social computing, personality, emergency response

## OPEN ACCESS

### Edited by:

Jie Yang,  
Delft University of Technology,  
Netherlands

### Reviewed by:

Dallia Durães,  
University of Minho, Portugal  
Thomas Mandl,  
University of Hildesheim, Germany

### \*Correspondence:

Federica Lucia Vinella  
f.l.vinella@uu.nl

### Specialty section:

This article was submitted to  
Machine Learning and Artificial  
Intelligence,  
a section of the journal  
Frontiers in Artificial Intelligence

**Received:** 19 November 2021

**Accepted:** 05 May 2022

**Published:** 27 May 2022

### Citation:

Vinella FL, Odo C, Lykourantzou I and  
Masthoff J (2022) How Personality  
and Communication Patterns Affect  
Online *ad-hoc* Teams Under Pressure  
Front. Artif. Intell. 5:818491.  
doi: 10.3389/frai.2022.818491

## 1. INTRODUCTION

Situations that require working together, fast, and efficiently under pressure are on the rise, especially in an increasingly fragile global ecosystem (Schneider, 2011; Kretzschmar et al., 2022). From handling widespread geopolitical conflicts (Friede, 2022) to mitigating environmental disasters (Gay-Antaki, 2021), several organizations are investing in crowdsourcing intervention to aid large-scale mobilization of resources including emergency shelters and disaster-event detection (Pettet et al., 2022; Stephens and Robertson, 2022; Zhang, 2022). Likewise, virtual teamwork enacted in high-urgency, high-stress tasks is on demand. Grassroots social engagement [i.e., Covid-19 pandemic hackathons (Colovic et al., 2022)], incident response squads (Palen et al., 2007), community response teams, and on-call software solution teams (Anderson, 2020) are all examples of ongoing large-scale collaborative efforts. Emergency teams are devolving into technology, and the internet, in particular, to

enforce the timely resolution of complex problems within limited time frames, often under stress, and potentially with collaborators who have never worked together in the past. The benefits of working virtually and remotely are evident as shown by the thriving field of telemedicine with remote surgical teams aiding medical centers in coping with widespread pandemics (Etheridge et al., 2022). Nevertheless, little is known about the factors that can make or break such teams. In this study, we attempt to answer questions such as: *What are the personality characteristics that render high-stake online teams successful? Which skills, abilities, or socio-cultural elements are essential to consider while forming these teams? Are there any particular communication patterns that can serve as early signals of effective teamwork under stress?* Answering these questions is crucial to leverage available resources and intellect in critical situations. Although group research has since long investigated the effect of factors including personality, knowledge, skills, or socio-cultural facets on virtual teamwork (Kichuk and Wiesner, 1997; Krumm et al., 2016), few studies examine these characteristics on the specific problem of online collaboration strained by external—psychological or time-related—aspects.

Teams performing in rapid response environments do not perform similarly to “normal” teamwork settings. They are under pressure from the high-demand context under which they operate. The time-bounded nature of the task increases the chances of failure (Driskell et al., 2018). Characteristics of team performance in rapid-response, high-stress contexts are team members’ ability to work in a team and personality traits (McManus et al., 2004; Subramaniam et al., 2010). However, to date, studies on high-stake teams focus either on emergency professional teams, crowd participation in emergency response, or the collaboration between these two groups without considering the aspect of team formation at the crowd level. Our study observes **remote, ubiquitous, online, and *ad-hoc* crowd teams** instead of traditional emergency response offline teams with specialized individuals (Chen et al., 2008). We deem the crowd, alongside teamwork emergency response, as the two most relevant aspects of this research, as we analyze and report properties contributing to successful outcomes under situations of stress and ambiguity. Furthermore, we examine the relationship between personality, socio-cultural elements, and communication patterns on the one hand, with team performance and satisfaction on the other, in the context of *ad-hoc online teams in rapid-response, high-pressure tasks*.

## 1.1. The Task: A Virtual Maze for Remote Crowdsourcing Emergency Teamwork

To study participant interactions in *ad-hoc* teams of strangers under pressure, we turn to crowdsourcing, and a custom-made task. Our task is inspired by the “Keep Talking Nobody Explodes” (Knuth, 2021) puzzle video game. Participants work in dyads, and their common mission is to defuse a bomb that is placed within a maze, by combining information that is unique to each one of them. One participant is assigned the role of the “Defuser”: they can “walk” inside the maze toward the bomb and defuse it, but they do not know where the maze walls are. The other participant

is assigned the role of the “Lead Expert”: they have the map of the maze but they cannot walk in it. The Defuser and the Lead Expert must exchange information and actions, to defuse the bomb within a limited amount of time. The task has been designed to have the same critical characteristics as actual emergency response tasks, namely a high-demanding environment, enforced role division, performance pressure and stress.

### 1.1.1. High-Demanding Environment

Instances of crisis constitute a large part of what emergency teams have to deal with and radically define their functional and structural properties. Demanding environments have critical requirements with tangible consequences for poor performance (e.g., accidents, errors, stress). By portraying the element of urgency in the form of a virtual bomb and increased time pressure (Bell et al., 2018) we focus on a single objective—reaching the bomb on time—and deliver the results of a study task that is critically cooperative and built for productive communication. In our setting, virtual crowd teams must deliver innovative solutions and deliver them quickly. The typical environmental constraints of high-demanding tasks (time, urgency, risks) command for independent, stable, role-defined teams sharing mutual trust, values, and focus. As we reduce and inter-mediate communication through digital means, we impose an even further reliance on mutual objectives, well-defined roles and obligations, effective communication, and commitment.

### 1.1.2. Enforced Role Division

During cases of emergency, each team member has a distinct and specific role to play (Baldwin and Woods, 1994), which is typically a-priori and externally defined. Emergency and periods of crisis often create the need for established protocols of interaction respective to each part (Harrison and Connors, 1984). Although role division is typically fixed for these response units (e.g., medical, logistic, security, public relations, etc.), it must nonetheless be adaptable when facing unpredictable outcomes. By assigning strangers to pre-defined roles, we replicate a scenario where team roles are agreed upon yet flexible and interposed. Through well-defined roles and responsibilities, we evaluate the matching capabilities of crowd workers and investigate what are the constituents that fundamentally determine the execution of role-based virtual teamwork emergency response.

### 1.1.3. Performance Pressure and Stress

Prior work has shown that users involved in games such as the crowdsourcing task exhibit various forms of stress (Sabo and Rajčáni, 2017) and heightened emotional states (Hart et al., 2018). These teams are more susceptible to allostatic load, i.e., the process of “wear and tear” experienced by team players facing stressful conditions (Davaslioglu et al., 2019). Regarding the definition of stress, there are two kinds of stressful conditions and stressors (Ma et al., 2021). One definition follows the general assumption that a stressor (the triggering factor) negatively affects the person by degrading performance; the other sees stress as a challenge that improves performance and individual

**TABLE 1** | Positive and negative facets of the BIG-5 personality traits (Neuman et al., 1999).

Big five traits	Positive facets	Negative facets
Extraversion	Social, talkative, assertive, active	Retiring, sober, reserved, cautious
Agreeableness	Good-natured, gentle, Cooperative, hopeful	irritable, suspicious, uncooperative, inflexible
Conscientiousness	Self-disciplined, responsible, Organized, scrupulous	lacking self-discipline, irresponsible, Disorganized, unscrupulous
Emotional stability	Calm, enthusiastic, Poised, secure	Anxious, depressed, Emotional, insecure
Openness to experience	Imaginative, sensitive Intellectual, curious	down-to-earth, insensitive, simple, narrow

gains (Zhang and Lu, 2009). In this research, we stripped the task from several elements of the original video game with the intent to transverse from multiple sources of hindering stressors [that increase environmental demands and exceed the available resources (Salas et al., 1996; Gardner, 2012)] to a unique challenge to inspire and motivate collaborators. Finally, virtual teams experience stress differently than offline ones as they tend to experience lessened social support (Su et al., 2012) which exacerbates predispositions to stress and anxiety (Tarafdar and Stich, 2021). For this reason, even though we adjusted the task to limit encumbrance, we still regard the individual and team response to a stressful task as the determining factor for whether personal characteristics and/or team compositions help handle the challenge successfully.

By engaging the players in this high-pressure challenge, we examine whether personality characteristics (Conscientiousness, Extraversion, Neuroticism, Agreeableness, and Openness) may make individuals more prone to cooperation under time pressure. We further evaluate which, if any, combination of personalities results in better than average team performance. Similarly, we examine whether additional factors such as the participants' socio-cultural background affect their actual ability to work together and their satisfaction with teamwork. Understanding the crowds perception of the collaboration (and not only performance) will help the development of AI agents to support their needs—and not only effectiveness—in times of crisis. Additionally, perceptions on the collaboration may provide insights into why certain teams are more effective than others, and what teams may be willing to work together again on the next task. Thanks to the heterogeneous data gathered during the experiment, we look at the dyadic communication to unravel indicators of a given team's potential to cope with a high-demanding task under time pressure.

A focus of this research is the impact of participants' personality on *ad-hoc* online teamwork, that is crowd-sourced, brief, and under pressure. We use the Big Five personality model (Goldberg, 1990), also known as the Five-Factor model, to model and comprehend the relationship between crowd workers' personality traits and their disposition for online teamwork in emergency contingencies. We selected the Big Five model as it

is most commonly used for personality analysis [e.g., Highhouse et al., 2022; Ikizer et al., 2022; Mammadov, 2022] and for artificial intelligence systems that automatically adapt to personality [see (Smith et al., 2019) for a review of personality models used for personalization in persuasive technology, intelligent tutoring systems and recommender systems]. Additionally, many validated instruments exist to measure the Big Five traits, including the brief version of the Big Five Personality Inventory (Rammstedt and John, 2007) which we use in this paper. The Big Five model distinguishes between 5 traits<sup>1</sup>, each of which has multiple facets (see Table 1)

## 1.2. Research Scope: Human Factors for AI Intervention in Crowdsourcing Emergency Response Teams

As work shifts to increasingly digitized spaces and connections between collaborators are made broader by mobile and ubiquitous computing, we consider evaluating ways to channel these resources to help remote, crowdsourced emergency teams. Identifying attributes and interactions used in emergency crises can help organizations and research improve upon methods for remote communication. Our knowledge of characteristics that contribute to virtual emergency response teamwork can inform artificial intelligent systems in assessing whether and how an individual can be part of a response unit with limited time and resources, and also, if multiple possible workers and tasks exist, who to use for the emergency response teams.

The rest of the paper is organized as follows. Section 2 presents and discusses related work, including an overview of traditional teams under pressure and crowdsourcing efforts in this domain, as well as the study hypotheses. Section 3 describes the study design, including participant sample and task design. Section 4 describes the metrics used to capture participants' demographic characteristics, Big Five personality traits, and ability (prior experience and self-perceived ability), as well as the metrics of teamwork, namely: collaboration quality and communication patterns. Section 5 presents the results. In Section 6 we discuss the implications of this work, its limitations, and possible extensions for the future. Finally, section 7 concludes the paper with key findings and closing remarks.

## 2. RELATED WORK

### 2.1. Teams in Classical High-Demand, Time-Pressing Settings

#### 2.1.1. Operational Setting and Problem Scope

Significant research effort has been placed over the years on teams that need to perform in situations that require spontaneous, *ad-hoc* decisions and short-term planning, to resolve ambiguous or uncertain events, and where the consequences of failure are significant (Reuter et al., 2014). The scope of the problems that such teams are called to deal with is broad. It can include responding to natural disasters, like floods, hurricanes, and fires, but also managing crises (King, 2002), such as

<sup>1</sup>Emotional Stability is often replaced in literature by its opposite Neuroticism.

terrorism events (Longstaff and Yang, 2008), events occurring in long-duration spaceflights (Salas et al., 2015), nuclear plant control rooms (Stachowski et al., 2009), or situations taking place in a military context (Driskell et al., 2014). It can also include more benign everyday workplace settings, such as on-call software teams dealing with organizational incidents, like security or service failure events (for example the recent Google outage (Bergen, 2020), journalist teams for the immediate coverage of unexpected events (Archibold, 2003), but also short-term project teams (Galbraith and Lawler, 1993) and task forces (Hackman, 1990). Their size can vary, from dyads and triads (Foushee, 1984), to dozens (Helmreich, 1967), to twenty or more (Stuster, 2011).

### 2.1.2. Differences From Normal Teams

What separates these teams from teams in “normal” settings, is the extreme, atypical environment within which they operate, which overall entails time pressure, high levels of risk, increased consequences for poor performance (Driskell et al., 2018), no previous work experience with one another, and the need to perform their task almost immediately on team formation (Mckinney et al., 2005; Mendonça, 2007). Harrison and Connors (1984) use the term exotic environment to describe a work setting that is marked by hostile environmental demands, restricted working conditions, isolation from those outside the setting, and confinement and enforced interactions for those inside it. Using the related term extreme environment, Bell et al. (2018) add that these settings are also characterized by limited time to finish the task. Performance pressure and severe consequences for ineffective performance are also characteristic of these settings, and this pressure can act as a double-edged sword that can lead the team to outstanding performance, or cripple it Gardner (2012). The tasks that teams in these settings must solve are usually characterized by ambiguity and urgency (Yu et al., 2008; Stachowski et al., 2009).

### 2.1.3. Factors Affecting the Success of Emergency Teams

Which factors determine team success in this high-demand, high-stress environment? *Skill* and expertise are the primary factors. Teams traditionally trained as emergency response units rely on the specialized expertise of the stages of the incident response and carry insider knowledge of the organizational policies, their obligations, the communication channels, and the tools supplied by the hiring organization. Thereof, the effectiveness of traditionally formed emergency response teams relies to a great extent on the level of preparedness and competence of the hiring body (or authority) that trained and assembled them, with multiple historical incidents providing evidence for the need for precise training programs and hiring criteria (Alexander, 2003). Examining command and control teams, Ellis et al. (2005) find that team members with higher training demonstrated greater proficiency in planning and task coordination activities, as well as in collaborative problem-solving, and communication. The study also found that it is the knowledge competencies of the team member with the most critical position that benefited the team the most.

The second factor of interest is the allocation of *roles and authority*. A prominent characteristic of typical high-stake teams, such as STAs (swift-starting action teams), is that they comprise experts (Mckinney et al., 2005) with specific roles and responsibilities. Multiple studies confirm the value of stable role structure in the division of labor and in enhancing the predictability of team interactions, allowing each team member to know what to expect from their teammates in critical situations (Hackman and Morris, 1975; Stachowski et al., 2009). The reason is that misunderstandings or disagreements about authority and role accountability (especially non-desirable roles like clean-up) may lead to team conflict, especially in the presence of unprecedented emergency response tasks (Quarantelli, 1988; Weick, 1993). The meta-analysis of De Wit et al. (2012) further confirms the negative relationships between process and role conflict, and team results such as cohesion, commitment, and performance. On the other hand, flexibility, the ability to improvise, and entrusting functional requirements to determine roles, rather than relying on titles may also be of benefit (Briggs, 2005; Mendonça, 2007). A highly defined role structure with clear roles seems to benefit more tasks that are structured. On the contrary, a flatter structure may be better for ambiguous tasks for which no apparent solution can be easily found (Worchel and Shackelford, 1991) (such as the task of responding to the 2001 World Trade Center attack Mendonça, 2007).

*Personality* is another prominent factor affecting the success of high-stakes teams, in line with the broader personnel selection literature which indicates that if relevant personality factors are identified for a specific job, future performance can be predicted (Borman et al., 1980). Using the occupational personality questionnaire to study the emergency command ability of offshore installation managers, Flin and Slaven (1996) finds significant correlations between command abilities in critical situations and certain personality elements. From their results, it appears that the highest-rated performance came from those who (a) like to take charge and supervise others (high score on controlling), (b) consider themselves to be fun-loving, sociable, and humorous (high score on outgoing), (c) are less interested in analyzing human behavior (low score on behavioral), (d) are more interested in practical than abstract problem solving (low score on conceptual), and (e) prefer to make decisions quickly rather than take time to weigh up all the evidence (high score on decisive).

Flin and Slaven (1996) contribution, however modest in size, is only pertinent to emergency command responsibilities and applicable only within a specific type of organization (offshore installation managers). Other researchers have focused on the possible existence of a “rescue personality,” in multiple additional domains where emergency services and occupational stress are pivotal. Kennedy et al.’s (2014) research on how personality influences the workforce decisions of emergency nurses reveals that certain traits matter more than others. High Extraversion, Openness to experience, and Agreeableness were especially common amongst emergency nurses. Extraversion was also present among emergency department senior medical staff (Boyd and Brown, 2005) as part of the controversial

ENTJ (Extrovert, Intuitive, Thinking, Judging) personality type<sup>2</sup> (Myers, 1962).

Partially supporting these findings is the work of Wagner et al. (2009) on the personality traits of paid professional firefighters. Although high Conscientiousness was not a determinant factor in this vocational role, Extraversion had significance. Certain personality traits seem to cluster under particular types of emergency professions; the differentiation between correlation and causality between these two variables is not always easy to untangle. Feelings of anxiety and insecurity, as well as heightened levels of Neuroticism and Openness, were seen to be most likely the results, and not the cause, of the repetitive exposure to experiences of loss and distress (Pajonk et al., 2011). By broadening the sample to the general public (virtual crowd), we aim at decoupling the effects that a specialized profession could have on one's propensity to emergency response.

Finally, certain *interaction patterns* are useful predictors of whether an *ad-hoc* team that has been brought together for immediate task performance will succeed or not, in classical emergency response teams. Although swift-start teams have little time to build their group processes before starting to work on the task, it is also known that team routines get established early in the team's lifecycle. The same initial interactions have an effect on subsequent communication and norms (Gersick and Hackman, 1990). The study of Zijlstra et al. (2012) reveals that there are certain early patterns of communication that distinguish effective from less effective teams. Specifically, they find that effective teams engage in communication that is more stable in duration and complexity, more balanced, and less monopolized by a single participant compared to inefficient teams that exhibit frequent mono-actor patterns, consisting of a single team member posing and answering their questions and commenting on their observations. They also found that efficient teams exhibit more reciprocity and trust, with the team members engaged and in the same direction of action toward the task goal. The presence of trust as a crucial factor is also highlighted (Wildman et al., 2012). The study of Waller et al. (2004) reveals that efficient teams in non-routine situations focused their actions on information collection and task prioritization. Finally, Kanki et al. (1991, 1989) complement the above by showing that the communication of effective swift-start two-person crews focuses on immediate task execution, expressed as low-complexity, straightforward action statements, and is less focused on other non-standard communication.

Although classical rapid-action teams are widely studied, these literature findings do not necessarily translate to online crowd rapid-action teams. Traditional emergency teams comprise highly trained professionals with a shared understanding of the crisis domain, and often a shared loyalty to an organization. In contrast, crowd teams mainly consist of non-experts, and they are more volatile and heterogeneous regarding the motivators

that draw their members to the particular task. Considering the multiplication and globalization of the events that require swift action, it is likely that in the future, we will need to turn more and more to crowd workers and volunteers to form *ad-hoc* online teams that can deal with high-stake situations under pressure. In this light, the extensive study of classical rapid-action teams can provide us with the first grounded indications of specific parameters to look at to identify predictors of successful team formation in online crowd action teams. Given that in a crowd setting, the allocation of roles is likely to take place based on arrival and availability, in this work, we focus on the parameters of personality and communication patterns as predictors of forming a successful crowd team to tackle unforeseen situations under time pressure.

#### 2.1.4. Onsite and Offsite Emergency Response Teams

The history of emergency response teams—and more broadly of emergency preparedness—is essentially as old as societal and humanitarian threats. For as long as emergencies have affected human lives, societies have found collective ways to organize efforts to mitigate, prepare, respond, and recover from the aftermaths of crises. Emergency preparedness programs have evolved along with societal changes and technological advancements. Notable historical events such as the first world war brought national societies to unify and strengthen their approaches to natural, intentional, and accidental disasters (Herstein et al., 2021). The International Federation of Red Cross and Red Crescent Societies is one of the most prominent products of global pursuits unifying volunteer networks, community-based expertise, and independent advisers into standardized practices (London, 1998). As emergency response evolves, emergency response teams reshape ways to communicate and function in an era of accelerated technological progress.

Formerly, emergency teams operated face-to-face and on-site in response to environmental disasters (Brennan and Flint, 2007), war conflicts (Abdul-Razik et al., 2021), and epidemics (Leach et al., 2022). With the broadening digitization of services, society is increasingly reliant on technology for its functioning. The so-called information era entails the vast market of the internet of things, software, and the worldwide web to enable widespread financial and data transactions (Stehr, 2001). Technological dependency is making us faster and smarter and, at the same time, more vulnerable to novel threats (e.g., malware attacks, identity theft, financial fraud, security breaches, etc.). Emergency response teams not only must face novel and extensive digital threats but must also learn to leverage the resourcefulness of recent technology [ubiquitous computing (Smirnov et al., 2011), robotics (Kawatsuma et al., 2012), simulations (Kincaid et al., 2003), smart sensors (Abu-Elkheir et al., 2016), and social media networks Potts, 2013] to strengthen their outreach and preparedness.

Overall, the vast majority of emergency response teams operate in a hybrid fashion combining onsite support with online offsite communication. Some others divide efforts between online and face-to-face tasks depending on the phase of the response (i.e., mitigation, preparedness, response, and recovery Brennan and Flint, 2007). Relevant to our research is the

<sup>2</sup>studies have been conducted on construct MBTI validity and test-retest reliability (including a meta-study by Capraro and Capraro (2002) which showed good results), others have argued that there are scientific limitations to these studies, the use of MBTI, and its underlying theory (e.g., Boyle, 1995; Pittenger, 2005; Stein and Swan, 2019).

pertinence of virtual communication channels in the large-scale crowdsourced emergency response domain that is typically remote, collaborative, and online. To define our target group, we firstly identify general characteristics that, in the classical sense, differentiate between onsite and offsite emergency response teams. Although the two domains share very similar objectives and attributes such as organizational culture, expertise, team structure, communication, and teamwork (Leach and Mayo, 2013), since their capabilities and duties differ, some of these attributes are more imperative than others. In the following subsections, we introduce two representative attributes critical for each teamwork domain.

#### 2.1.4.1. Onsite Emergency Response Teams

Two prominent attributes of onsite teams are **experience** and **coordination**. Teams working onsite are usually part of rescue operations (Chen and Miller-Hooks, 2012) and disaster relief (Bjerge et al., 2016) that require the participation and coordination of experts. These include fire and rescue services and police forces, commercial entities, volunteer organizations such as the Red Cross, media organizations, and the public (Yang et al., 2009). The need for distinct expertise requires teams to develop and apply specialized knowledge. Onsite emergency response experts can hold intelligence on chemical properties, procedures for reporting emergencies, fire and protective equipment, decontamination, and evacuation gained through training, experience, and/or formal education.

Without qualified knowledge and standardized procedures, onsite emergency response teams would fall short of promptly and accurately addressing ongoing crises. Equally important is coordination among experts as onsite emergency must successfully distribute superintendence and responsibilities between diverse professionals for effective prevention, preparedness, and response to emergencies. In their work on coordination in emergency response management, Chen et al. (2008) developed a life-cycle approach with three distinct sets of activities on the timeline continuum (pre-incident phase, during incident phase, and recovery phase). The cycle closes after de-briefing and when actionable items are learned from the intervention and incorporated into the plan to affect future preparedness (Chen et al., 2008). The same authors identified several elements of coordination such as activities, coordination objects, and constraints that differ between phases and between cultural, political, regulatory, and infrastructural properties of emergency response.

#### 2.1.4.2. Offsite Emergency Response Teams

Two distinguishing attributes of offsite remote emergency response teams are **communication** and **sensemaking**. While onsite teams converge in rescue operations and disaster relief, remote offsite emergency response teams outreach and distribute resources. Known crises overseen by offsite emergency response teams are air-traffic control (Hughes et al., 1992), subway crisis management (Heath and Luff, 1992), and emergency response call centers (Normark, 2002; Pettersson et al., 2004). Although clear roles are important in these teams, clear communication is of the essence. Depending on the kind of interaction (e.g.,

serendipitous, inbound, and outbound Landgren and Nulden, 2007), and the referent (e.g., non-experts' communication, situation update, situational awareness, services access assistance Velev and Zlateva, 2012), clear communication and interaction protocols fundamentally determine the interaction mediated by computer systems for offsite rescue teams.

Through clear communication, offsite emergency response teams can harvest sensemaking. This is the collection of actions that make the situation understandable and that prevent an escalation of the emergency (Landgren and Nulden, 2007). Sensemaking has properties such as identity construction, retrospection, enactment, social reactions, dynamism, environmental cues, and plausibility (Muhren et al., 2010). The importance of sensemaking in a remote emergency context is ever so apparent due to the practical constraints that teams experience as they communicate remotely. According to Weick (1993), most shortcomings from failed emergency responses are due to a deficiency in sensemaking (or contextual rationality). Weick (1993)'s work uncovers four potential sources of resilience that make *ad-hoc* groups less vulnerable to disruption of sensemaking. These sources are (i) improvisation, (ii) virtual role systems, (iii) the attitude of wisdom, and (iv) norms of respectful interaction. Weick (1993) analyses the dynamics of role structure and sense-making occurring in the historical Mann Gluch disaster. The incident served as an example of what needs to be re-examined about temporary systems, structuration, non-disclosed intimacy, inter-group dynamics, and team building (Weick, 1993), especially important for offsite emergency response operations.

The design of computer-mediated emergency response also needs to be informed by an understanding of the cognitive processes involved in responding to unanticipated contingencies (Mendonça, 2007). These cognitive factors, defined by Mendonça (2007), are directly linked to the specificity of emergence management and its characteristics of rarity, time pressure, uncertainty, high and broad consequences, complexity, and multiple decision making. Besides, computer-mediated emergency response teams are much more predisposed to incorporate the output of citizen convergence (Schmidt et al., 2018) into their work than traditional onsite rescue teams. However, as developments in online informational convergence change the remote domain of rescue operations, citizens and crowds are bringing novel paradigms. These include unfamiliar team members, ill-defined tasks, fleeting membership, multiple and conflicting goals, and geographically distributed collaboration (Majchrzak and More, 2011). In the following section, we explore the topic of crowdsourcing for emergency response.

## 2.2. Crowdsourcing for Emergency Response

### 2.2.1. Emergency Response Through Individual Crowd Contributions

Crowds are increasingly involved in response to emergencies. The characteristic of emergency response crowdsourcing is the short-lived engagement in the task. Crowds' contributions consist

of primarily individual, one-time, and remote interactions. This “long-tail” of contributions is a well-observed phenomenon in most content-oriented online communities (Shirky, 2008). The role of these one-time crowd users is important when it acts as a fast and ubiquitous response to urgent, environmental and social crises (hurricanes, terrorist attacks, widespread fires, large oil spills, etc.) (Heinzelman and Waters, 2010; Yuan and Liu, 2018; Chau, 2020), protest movements (Elsafoury, 2020), but also activism (Farkas and Neumayer, 2017; Lee, 2020) and civic participation (Hemphill and Roback, 2014; Mitchell and Lim, 2018). In critical scenarios of this kind, the crowd is intended as a manifold social tool by servicing as a reporter, social computer, sensor, and executor of both micro and macro-tasks.

Several theoretical studies propose system models and features designed to facilitate the positioning of the crowd as the leading resource for emergency management. In the domain of communication technologies for health care Hossain et al. (2017) suggest benefiting from the users’ social contacts to trigger a faster response, or to make the most of crowdsourcing attributes—such as collaboration and tournaments—to attract the right crowd for the job. From a complex systems perspective, Song et al. (2020) propose harnessing the self-organizing operation mechanisms of crowdsourcing for efficient disaster governance. In the context of natural disaster management, Ernst et al. (2017) propose hybrid systems that rely on the remote coordination of volunteers to collect location-dependent information, which in turn can support emergency managers making quick but solid decisions. Elsafoury (2020) propose another hybrid feature, this time combining machine learning with crowdsourcing to rapidly detect protest repression incidents through social media.

Specific crowdsourcing tools and platforms address emergencies. Poblet et al.’s (2013) review indicates that these platforms belong to two main categories, namely: (i) data-oriented, and (ii) communication-oriented. The first category concerns tools developed for the intensive aggregation, mining, and processing of data gathered through the crowd. The second category aims at supporting communication between crowd users and disaster management systems by allowing seamless interaction between them. The platform “Ushahidi” (Okolloh, 2009) is one example of a crowd application designed to decentralize the support of volunteers for the report of violence in Kenya, by collecting sensitive reports, organizing rapid response actions across multiple agencies, documenting ongoing changes, generating automatic alerts from user updates and visualizing data streams in real-time.

In another example, several digital volunteer organizations (Standby Task Force, Humanity Road, and Open Crisis) have integrated social media monitoring in their systems when cooperating with other humanitarian bodies in disaster relief operations (Poblet et al., 2013). Poblet et al.’s (2013) review of crowdsourcing tools for disaster management offers an extensive list of crowdsourcing tools, including online platforms and mobile applications across the globe. Aside from those tools that support response and recovery-based only efforts, others, such as ArcGIS (Allen, 2011), Sahana (Careem et al., 2006), OpenIR (Ducao, 2013), and CrisisTracker (Rogstadius et al., 2013), provide support for mitigation and crisis preparedness. These tools pivot around the crowd for achieving great humanistic

and environmental causes while leveraging the strength of geographically dispersed collaboration.

However, despite the growth of several initiatives and digital platforms designated to facilitate crowd intervention in emergency response, these initiatives are primarily based on individual contributions, without taking advantage of team dynamics that can arise among the crowd participants. This lack of communication, either due to team conflict (Yeo et al., 2018), or unfitness of the tools (Dilmaghani and Rao, 2006), makes crowdsourcing efforts less efficient, which often fail to address the event at hand, either as standalone initiatives or as supporting capacity to expert emergency management (Heath and Palenchar, 2000). Beyond the subject of crowdsourcing for emergency response, other team categories are also relevant to our research on *ad-hoc* crowd team formation. Action teams, rapid response teams, and citizen science, to name a few, are groups formed through the crowd and behave similarly to *ad-hoc* teams. Similar entities could benefit from system improvements addressing better team formation and communication strategies adopted from a better understanding of team dynamics in stressful situations. In the following subsection, we elaborate on existing—albeit early—efforts that seek to involve the crowd in formations and groups.

### 2.2.2. Crowd Cooperation for Emergency Response

Aside from individual crowd contributions, a few studies have looked into facilitating communication among crowd members to respond to and manage unexpected events. Providing people with communication channels can help them gain a broader view of the event they need to deal with (Perez and Zeadally, 2019), and better coordinate their efforts (Martella et al., 2017). Song et al. (2020) analyzed a total of twelve international case studies of crowdsourcing and natural disaster governance. They denote that, across all of these instances, the crowd manifested (at least at some level in their response mechanisms) self-organizing properties that lead its individuals to form collaborative ties spontaneously. It suggests that the multi-directional relationship between the crowdsourcing platforms, the initiators, and the contractors, while not strictly guided, triggers the formation of functional teams that act as active response units. Under this instance, the crowd forms *ad-hoc* groups as the emerging outcome of community disaster resilience (Song et al., 2020). As long as collaboration is advantageous in emergency response and time management remains vital in real-life crises, boosting the efficacy of crowd participation starting from the level of team formation can get teams closer to their desired outcomes.

Many combinations of individual traits add up as building blocks for the entire social entity that is the team. Assuming that the single characteristic is, at least in principle, an optimal fit for the task, the way it interacts with the rest of the teammates’ features is equally relevant. Personality clashes are present in virtual team interactions just as in traditional face-to-face cases. Following Van de Ven et al. (1976) definition of teams as “groups becoming more effective over time,” Salehi et al.’s (2017) work on stable crowd teams recognizes familiarity as the utmost important factor that enhances team performance. However, familiarity is a variable that cannot always be factored in when teaming up with individuals part of a virtual crowd, who are

often sporadic contributors. Therefore, while familiarity in crowd teams has its tangible benefits (Salehi et al., 2017) for more stable tasks (like creative ones), relying on team familiarity to form effective crowd teams is not always feasible for short-lived, unpredictable, and mutable tasks.

For relatively short-lived assignments, the distribution of personality types matters more for the success and the establishment of trust in crowd teams than the pervasiveness of one specific type. Lykourantzou et al.'s (2016) work on crowd teams shows that balancing personality traits not only leads to significantly better performance on collaborative tasks but also reduces conflict and heightens the levels of satisfaction and acceptance. Holistically, when considering the impact of personality distribution in crowd teams, aspects other than personality traits play an often overlooked yet fundamental role. As Lykourantzou et al.'s (2016) noted: *test Personality could also be examined with regards to task type. For example, competitive tasks (like ideation contests among competing crowd teams) may amplify clashes within imbalanced teams, more than collaborative tasks.*" We aim to uncover the relevance of personality, communication, and other factors in a virtual emergency response task. Unlike other studies (Floch et al., 2012; Vivacqua and Borges, 2012; Ernst et al., 2017) evaluating crowd emergency response as a collective and self-organized effort, we propose a team-specific approach to the formation of crowd emergency units that strongly connects with theories and models of teams composition, and assembly and team science (National Research Council, 2015).

Closing, most crowdsourced initiatives for high-stake, high-pressure tasks rely on individual contributions. Few works use some form of teamwork to coordinate crowd participants' efforts spontaneously and not according to a systematic approach or criteria. The formation of crowd emergency teams according to a set of characteristics with known expected effects could help these teams experience less interpersonal conflicts, establish team cohesion faster, and increase the teams' chances of success. In this work, we systematize online team formation for high-pressure tasks. We closely investigate the effects of personality and communication patterns, contributing to such teams' success and helping harness the crowd's potential better in emergency response.

### 3. STUDY DESIGN

Many factors may impact whether teams collaborate well and achieve their goals in an emergency response task. These include the demographics and personality of team members (both at the level of individuals and aggregated over the team), and the communication patterns used. This study explored which factors matter for team success and perceptions of collaboration quality. Given the many factors and output measures considered, the study was exploratory in nature, with the aim to gain initial insights into what matters and in which way, to be tested further in follow on studies.

#### 3.1. Sample

120 Amazon Mechanical Turk workers (41 female, 78 male, 1 prefer not to say) participated. The task duration was

approximately 20 min. Most participants were of U.S. (67 users) and Indian nationalities (51 users), one participant was Irish and another one was British. The majority had College (87) or Postgraduate degrees (15), while some had either some college education (9) or High School (9). Most were between 30 and 49 years of age. For an overview of the demographic data of the sample see **Table 7**.

#### 3.2. Compensation

The participants received a base reward of \$3, and a bonus reward of \$3 if the challenge was completed successfully. The base pay was based on current fair crowd work compensation practices, whereas the bonus pay matched the base pay to double the reward for those teams that defused the bomb on time. The payment was weighted against the hourly rate or AMT workers as reported in Hara et al. (2018). In selecting the payment amount, we took into account three considerations from the literature (Olson and Kellogg, 2014; Lykourantzou et al., 2016). First, the payment had to conform to the community standards of the crowdsourcing platform so as not to bias the quality through workers who would accept low wages or workers who would only choose the task purely for its high compensation. Second, this payment had to cover the task duration. Thirdly, it took into account the demographics of the target worker population (minimum wage).

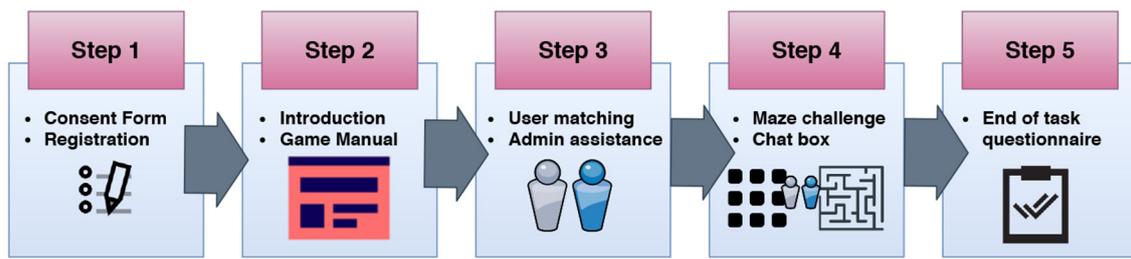
We recruited through the Amazon Mechanical Turk (AMT) Human Intelligent Task (HIT) platform. AMT was chosen for its breadth of crowd workers and its abundant labor availability, which is estimated to be no less than 2K workers at any given time, and over 100K workers overall (Difallah et al., 2018)<sup>3</sup>. No pre-selection was required to participate in the task. We intended to attract a large variety of participants, regardless of differences in background. The absence of pre-selection criteria may have influenced participants' written English, a limitation discussed in Section 6.2.3. Finally, the HIT itself contained information about the reward, the duration of the task, and a short description of the cooperative game.

#### 3.3. Task Design and Setting

Although the task was artificial it was designed as an analog setting enacting the key characteristics of the high-demand, high-pressure environments that we are interested in. These include:

1. **Simulated element of physical danger.** The consequence of the team failing to navigate the maze is a bomb exploding. Although participants were aware that they are playing a game, the element of physical danger, even an enacted one, alters their perception, with possible effects on the way they process information, coordinate their efforts, and discuss (Kamphuis et al., 2011).
2. **Pre-determined team roles.** The presence of these roles enables stable and predictable group interactions (McMichael et al., 1999) instead of relying upon the slower and

<sup>3</sup>AMT worker's population is composed primarily of Indian and American nationalities, followed by Chinese, British, and Philippino (Difallah et al., 2018). The gender is slightly predominantly female within the American sample and more male in other countries (Difallah et al., 2018). Its population average age is less than the world population average, as most AMT workers were born after the 1990's (Difallah et al., 2018).



**FIGURE 1** | System overview with the five steps of the study design. After registration, users arrive at an introductory page with relevant information about the task, and then they are matched in dyads on a first-in-first-out basis. Each team then proceeds to their dedicated virtual room where they cooperate to defuse the bomb in the maze within a given time frame. Finally, they fill out a questionnaire about their abilities and perceived collaboration quality.

autonomous differentiation of team roles (Belbin, 2012), which cannot always happen in circumstances of emergency. Predefined role-playing exercised control over one's limited access to information, which symbolizes the relationship between an overseeing entity (in our case, the Lead Expert) and an operative agent (in our case, the Defuser). Furthermore, similar to real-life action teams, team membership symbolizes work shifts (Zijlstra et al., 2012). It represents the random assignment of roles on a first-come-first-served basis. Similar to emergency response teams, this approach creates teams with little time to explore personal similarities and differences or to go through classical team development processes (Tuckman and Jensen, 1977; Lacoursiere, 1980).

3. **Stress and increased consequences of failure.** The novelty of the task, alongside its short duration, positions the crowd participants in a situation similar to emergency management scenarios. Here, the users need to act decisively within tight time schedules, often only with access to incomplete or difficult to decode information (Carver and Turoff, 2007). It means that the participants (a) absorb information rapidly, (b) judge by doing, (c) decide on the spot, (d) deal with the event with little preparation. Users are aware that their actions, if wrong, will cost them (and their teammate) reasonably significant retribution (in this case monetary) (Driskell et al., 2018). The combination of elements, namely: high-stake, time-constrained, fractional information, and role inter-dependency, makes this particular task a reasonably stressful one. More so, the original game "Keep Talking Nobody Explodes" has been utilized as a tool by past research to assess the effects of realistic stress on behavioral and physiological responses of participants (Sabo and Rajčani, 2017; Lee and Jung, 2020). These studies confirm that controlled environments of this sort can correctly reproduce similar stress levels of more realistic scenarios, thus inducing stimulus-response events—such as temporary homeostatic changes and speech variations— that signal increased stress.

To support the task setting, we designed a custom-made web system. The system pipeline, illustrated in **Figure 1**, was designed according to the following steps:

**Step 1: Consent form and registration.** Participants registered with a username, AMT IDs (unique identifier needed to reward them at the end of the task), demographic information (gender, age, nationality, and education level), and Big-Five personality traits (Table 3). By registering, the participants agreed with the terms of service and gave their informed consent.

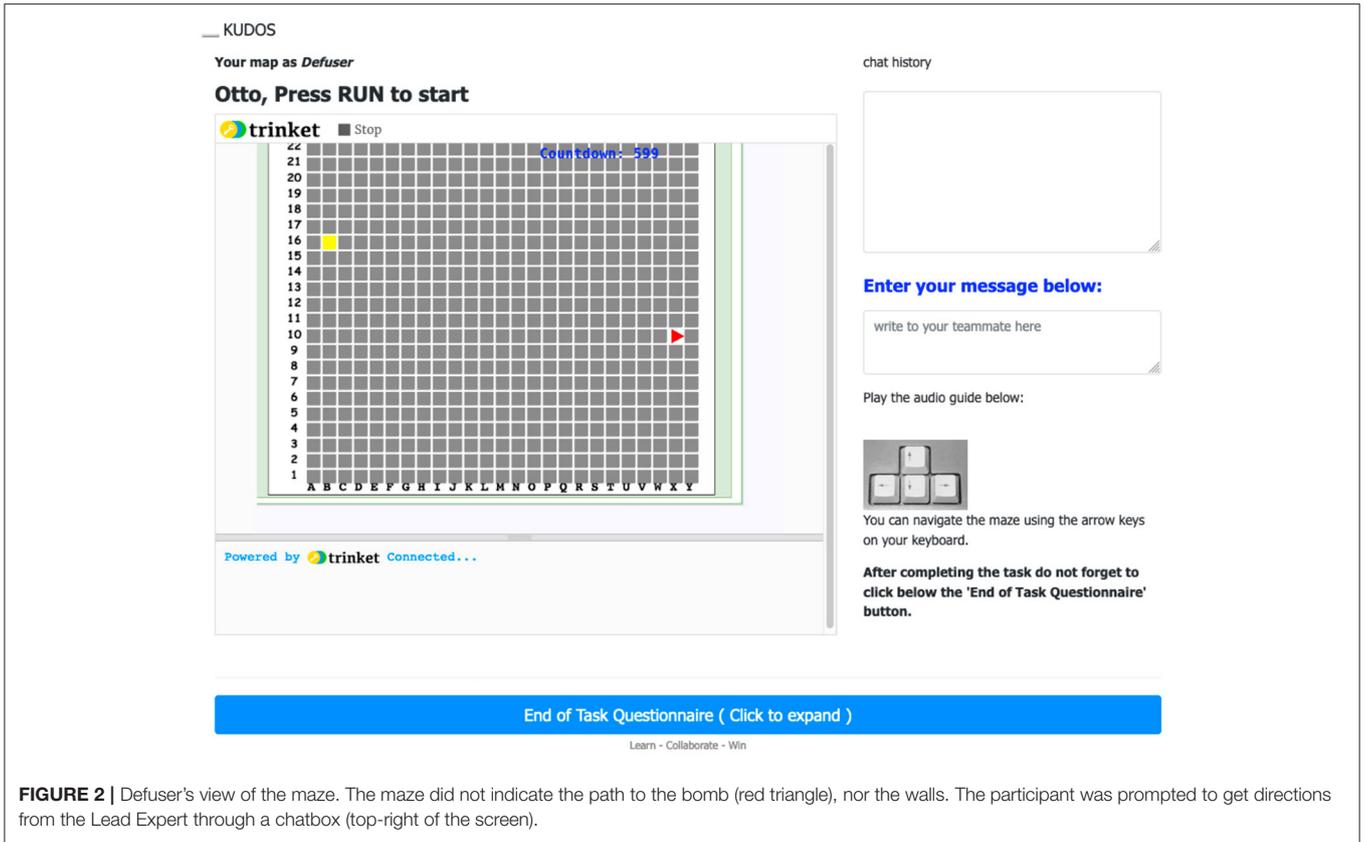
**Step 2: Introduction and game instructions.** After logging in, the "dangerous and challenging world of bomb defusing" (Knuth, 2021), the introductory page offered example screenshots of the two roles, instructions about the gameplay, plus information about the countdown and the end-of-task survey. The short info gave participants a broad idea of the task and focused on the platform functionalities (e.g., chat, game console, manual instructions, etc.).

**Step 3: User matching and admin assistance.** Participants entered the waiting room (i.e., matchmaking room) and were personally greeted by the system administrator while waiting for their teammates to join. If no other participants were present, they waited until a match would become available. The administrator also served as moderator and user support. The system allocated participants to teams in a first-in-first-out (FIFO) manner. As soon as two participants were present in the matchmaking room, they were placed together and asked to proceed to the main task (after first answering any questions they may have had).

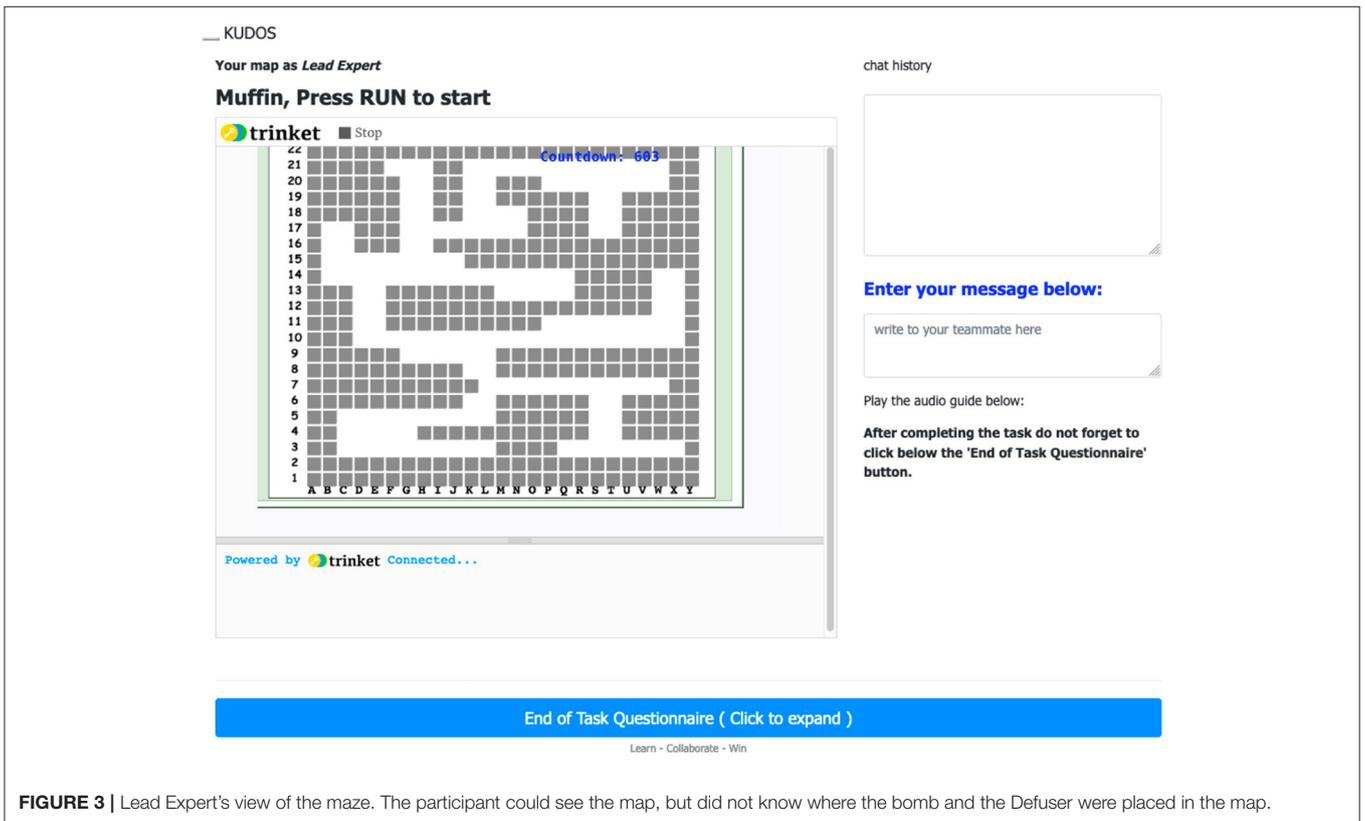
**Step 4: Maze challenge and chat box.** After matching, participants joined a private virtual room where they could see the maze game and chat to communicate with one another.

**Figure 2** shows what the Defuser saw. On the left-hand side, the Defuser saw a blind maze with their position (yellow square) and the bomb (red triangle). They could not see the walls as only the Lead Expert saw them. On the right-hand side, the Defuser saw the chatbox and, below it, a reminder to use the arrow keys to navigate the maze. Upon finishing the task, the blue bar at the bottom of the screen would take them to the final questionnaire.

**Figure 3** shows what the Lead Expert saw. The Lead Expert's view of the maze differed from that of the Defuser: they saw only the walls of the maze (gray squares) and the path to the bomb (white sections). The Lead Expert could neither see the Defuser in the maze nor the bomb. Both the Lead Expert and Defuser could see the same countdown and Cartesian coordinates of the maze, as well as the chatbox and the link to the final questionnaire.



**FIGURE 2** | Defuser's view of the maze. The maze did not indicate the path to the bomb (red triangle), nor the walls. The participant was prompted to get directions from the Lead Expert through a chatbox (top-right of the screen).



**FIGURE 3** | Lead Expert's view of the maze. The participant could see the map, but did not know where the bomb and the Defuser were placed in the map.

**TABLE 2** | Summary of variables.

	Variable	Type	Range	
Input	Personality <sup>5</sup>	Extraversion	Interval	2–10
		Agreeableness	Interval	2–10
		Conscientiousness	Interval	2–10
		Emotional stability	Interval	2–10
		Openness to experience	Interval	2–10
	Team Personality (for each trait)	StDev	Ratio	0–5.66
		Min	Interval	2–10
		Max	Interval	2–10
	Demographics	Mean	Interval	2–10
		Gender	Nominal	Male, Female, Other, not-disclosed
Age group		Ordinal	<20, 20–29, 30–39, 40–49, 50+	
Nationality <sup>6</sup>		Nominal	USA, India, UK, Ireland	
Education level		Ordinal	Less than High School, High School (HS), Some College (SC), College degree (Col), Postgraduate (PG)	
Communication patterns	Uncertainty, Action, Response, Planning, Factual, Non task-related	Ratio	≥0	
	Chat length (# Words)	Ratio	≥0	
	Chat total (# Posts)	Ratio	≥0	
Output	Performance	Nominal	Won, Lost	
	Perceived collaboration quality	Performance	Ordinal	1–5
		Cohesion	Ordinal	1–5
		Communication quality	Ordinal	1–5
		Balance	Ordinal	0–2
Satisfaction	Ordinal	0–2		

The Maze module was inspired by the video game “Keep Talking Nobody Explodes” (Knuth, 2021). It consisted of a 25 x 25 grid of squares with one square containing a yellow element (the position of the Defuser), one square containing a red triangle (the position of the bomb), and walls. Neither of the two players had access to all the information of the maze; they needed to cooperate. The Defuser could move inside the maze, by means of the four arrow keys, but they did not know where the walls were. The Lead Expert had the map, but could not navigate the maze. The Defuser’s role was to navigate the maze, with the help of the Lead expert, and defuse the bomb in time. Finally, a countdown timer was included, at the end of which the bomb exploded, unless it had been defused. The countdown started the moment both players entered the room. For this specific study, the timer was set to 400 s. After finishing the game, the participants received a validation code to submit to the AMT HIT for getting their base pay and bonus reward (for those teams that completed the challenge successfully). We deliberately excluded aspects of the original video game to reduce the number of variables and increase the controllability of the study environment. We wanted participants to focus on reaching the bomb on time without spreading themselves thin among the multi-modalities present in the original game (e.g., clues, strikes, wires, sequences, etc.). Besides, implementing most features of the original game would have added to the task complexity<sup>4</sup>. Hence, we did not include

<sup>4</sup>Also requiring considerably longer instructions and the introduction of manipulation checks to ensure instructions were read which further adds to task complexity.

penalties for the Defuser colliding with a wall. The only penalty—and end of game—was determined by the time running out before reaching the bomb. Furthermore, to ensure task brevity, we considered the bomb defused as soon as the Defuser stepped inside its cell. The simplification of the game has some limitations discussed in Section 6.2.

**Step 5: End of task questionnaire.** Participants rated the perceived collaboration quality on multiple aspects (see below), and also their abilities.

## 4. METRICS

We grouped the multilevel approach into two distinct classes referring to input and output variables (Table 2 provides a summary of all variables, their type and range.). Here the input metrics serve as the independent variables and the output ones as dependent variables.

### 4.1. Input Variables

#### 4.1.1. Big Five Personality Traits

To acquire a measure of the Big Five traits within the context of large-scale assessment under limited time and resources, we used the Big Five Inventory-10 (BFI-10) (Rammstedt and John, 2007). The inventory consists of ten questions (see Table 3).

<sup>5</sup>as the BFI-10 uses 5-point Likert scales one could argue that the data is ordinal, but given a total is calculated per trait we will regard it as interval.

<sup>6</sup>Free text entry, values provided here are those used by participants.

**TABLE 3** | BFI-10 instrument used, and its scoring: the trait for which each item was used and whether it was reverse scored (R)<sup>7</sup>.

I see myself as someone who ...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly	Trait	Reversed
1. ... is reserved	(1)	(2)	(3)	(4)	(5)	Extraversion	R
2. ... is generally trusting	(1)	(2)	(3)	(4)	(5)	Agreeableness	
3. ... tends to be lazy	(1)	(2)	(3)	(4)	(5)	Conscientiousness	R
4. ... is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)	Neuroticism	R
5. ... has few artistic interests	(1)	(2)	(3)	(4)	(5)	Openness to Experience	R
6. ... is outgoing, sociable	(1)	(2)	(3)	(4)	(5)	Extraversion	
7. ... tends to find faults with others	(1)	(2)	(3)	(4)	(5)	Agreeableness	R
8. ... does a thorough job	(1)	(2)	(3)	(4)	(5)	Conscientiousness	
9. ... gets nervous easily	(1)	(2)	(3)	(4)	(5)	Neuroticism	
10. ... has an active imagination	(1)	(2)	(3)	(4)	(5)	Openness to Experience	

Derived from the shortening of its lengthier predecessor (the Big Five Inventory (BFI-44) Rammstedt and John, 2007), it focuses on the psychometric characteristics of the BFI-44’s most representative items and reduces each Big Five dimension to 2 BFI items. The BFI-10 measures the personality traits of Extraversion, Agreeableness, Conscientiousness, Emotional Stability (Neuroticism), and Openness to experience (Rammstedt and John, 2007)<sup>8</sup>. For each trait, the BFI-10 score is calculated as the total score of the two statements associated with that trait, after reversing the score of some statements (see mapping of statements to traits and which statements’ scores are reversed in Table 3)<sup>9</sup>.

**4.1.2. Personality Traits of Groups**

There is no straightforward process for aggregating metrics such as personality traits for groups. However, the group recommender community has dealt with a similar issue namely the aggregation of group members preferences (Masthoff, 2004) and uses aggregation strategies from Social Choice Theory (Sen, 1986). Senot et al. (2010) distinguishes between (1) majority-based strategies that use the most popular values, (2) consensus-based strategies that consider the profiles of all group members, and (3) borderline strategies that only consider a subset. In our case, majority strategies do not apply given a group size of two. Of the consensus-based strategies, we use Average (which is also the most popular strategy in Group Recommender research). Of the borderline strategies, we use Minimum and Maximum<sup>10,11</sup>. Minimum is used as one may expect that team performance is strongly affected by the weakest member in the team, in line

with the popular saying “a chain is as strong as its weakest link”. Maximum is used as one may also expect that a strong member could make up for the weakness in another member (e.g., if one person is highly conscientious, they may entice the team to get the work done in time), particularly when the team is small. Finally, we used Standard Deviation (in line with the Cohesion metric introduced by Odo et al., 2019b), as the literature indicates the impact of diversity within teams<sup>12</sup>.

**4.1.3. Demographics**

Participants provided information about their gender, age group, nationality, and educational background. Socio-demographic measures identify characteristics that often influence the respondent’s opinions that could condition one’s behavior, culture, and experiences (Lavrakas, 2008). These socio-demographic factors provide further insight into the composition of teams, and what other characteristics—aside from personality traits—influence the collaboration. These socio-demographic factors that make someone distinct can turn into assets for group work. Therefore, by being aware of those characteristics, organizations and hiring bodies can better assemble and coordinate geographically dispersed teams (Muethel et al., 2012).

Multiple studies (Ruef et al., 2003; O’Leary and Mortensen, 2010; Akman et al., 2011) have identified various aspects of the teammates’ social backgrounds and demographic characteristics that condition teamwork. For example, members of similar demographic profiles have greater chances to kindle stronger affinity ties (Ruef et al., 2003). Other demographic differences, such as race, sex, age, and nationality, have also been found (Martins and Shalley, 2011) to affect the collective creativity of virtual teams. Age differences condition the creative processes of teams and intensify differences in technical experience (Martins and Shalley, 2011). Differences in nationality have a negative effect by interacting—however indirectly—with the technical experience of the teammates (Martins and Shalley, 2011).

<sup>7</sup>Reverse scored means that a 1 is changed into 5, 2 into 4, 4 into 2, and 5 into 1.  
<sup>8</sup>Test-retest correlations suggest acceptable reliability on a Likert scale of 1 (Disagree strongly) to 5 (Agree strongly). As prior studies have shown, the correlations of this instrument with other Big Five instruments, its correlations with self-and peer-ratings, and its associations with socio-demographic variables suggest good validity of the BFI-10 inventory (Rammstedt and John, 2007).  
<sup>9</sup>Reversed means that a score of 1 is changed into 5, 2 into 4, 4 into 2, and 5 into 1.  
<sup>10</sup>which in the Group Recommender community are called, respectively, Least Misery and Most Pleasure.  
<sup>11</sup>Personality traits likely differ on whether a high (or low) trait level positively or negatively impacts team performance. Using both minimum and maximum ensures this is no longer an issue.

<sup>12</sup>For teams of two, the use of standard deviation is equivalent to the use of numerical difference. We opted for standard deviation to build on the work by Odo et al. (2019b) and for generalizability to larger groups.

### 4.1.4. Communication Patterns

The methodology by Bowers et al. (1998) introduced a new approach to communication analysis prompted by a prior research gap in metrics that missed to analyze the more fine-grained interaction patterns other than simple frequency counts of words. They proposed the implementation of the categories of: (a) **uncertainty** statements, which included direct and indirect questions; (b) **action** statements, which required a particular member to perform a specific action; (c) **acknowledgments**, which were one-bit statements following uncertainty of action statements, such as "yes," "no," "roger"; (d) **responses**, which differed from acknowledgments only in that they conveyed more than one bit of information; (e) **planning** statements; (f) **factual** statements, which verbalized readily observable realities of the environment; and (g) **non task-related** statements. These categories quantified the performance of crews during simulated flight tasks, which improved the make-up of communication sequences analysis.

Based on Bowers et al. (1998) contribution, Davaslioglu et al. (2019) developed the Collective Allostatic Load Measurers system (CALM), which collected, aggregated, and analyzed data from individuals to make assessments on team situation awareness, performance, and resilience. The study used the virtual-reality game "Keep Talking Nobody Explodes" that we too used as inspiration for our experiments. Davaslioglu et al.'s (2019) study demonstrated that some teams exhibited patterns of communication, namely, action-response, uncertainty-response-action, and factual-uncertainty-response-action while working together under high-stress conditions. Acknowledgment statements, for instance, were seen to predominate more amongst high-performing teams, while low-performing teams had higher portions of non-task-related-statements. Similar studies on team communication analysis (Pfaff, 2012; Zijlstra et al., 2012) have identified patterns of communication. Given the proximity of our methodology to the studies of Bowers et al. (1998) and Davaslioglu et al. (2019), we implemented the same communication classes as they did. These communication patterns, or categories, are the following:

- **Uncertainty.** Uncertainty statements comprise questions (either direct or indirect) about the task (e.g., "Where are you at?," "Where is the bomb?").
- **Action.** Action statements indicate that one or both of the team members are taking action inside the game, or they are a direction to take action (e.g., "Move two steps down, then one right." "I am moving to position *x*," or "Go up for three blocks, then turn right").
- **Responses.** Response statements can accompany either uncertainty or action statements and suggest that a communication, or feedback loop (e.g., "yes," "no"), is ongoing.
- **Planning.** Planning statements that give the users a feeling that they are working together toward achieving a common goal. Planning statements can indicate the user's ability to reassess the situation, clarify the work, or plan the next actions.
- **Factual.** Factual statements are situational and describe the reality, for instance, by giving cues about how the maze looks

**TABLE 4 |** Example of an annotated chat sequence between a Lead Expert and a Defuser.

Text	Annotation	Role
Okay?	Response	Defuser
Got it?	Response	Lead Expert
I don't see bomb on my screen, do you know?	Uncertainty	Defuser
I'm the yellow square	Factual	Defuser
czzan't see bombs	Factual	Lead Expert
where r u?	Uncertainty	Lead Expert
16C	Factual	Defuser
go to 12x	Action	Lead Expert
where should I go?	Uncertainty	Defuser
One step at a time	Planning	Lead Expert
As a lead expert, I request you to guide me	Planning	Lead Expert
Both of us should use the code	Planning	Lead Expert
even I can't see the bomb	Factual	Lead Expert
there is a triangle on L3	Factual	Defuser
ok	Response	Lead Expert
wait	Action	Lead Expert
can you move? Take turns moving maybe?	Uncertainty	Defuser
follow my steps	Action	Lead Expert
How is your family members?	Non-Related	Defuser

like from the viewpoint of the Lead Expert, or at which coordinates the bomb is located.

- **Non task-related.** Non-task-related statements are parts of the chats that are categorized as non-related when they do not contribute to the achievement of the goal (e.g., "What is the weather like?").

**Table 4** illustrates an extract of the annotated chat between the Lead Expert and the Defuser. The patterns were labeled for each participant's text entry and annotated by two independent evaluators. The inter-rater agreement of the annotation was sufficiently high to be utilized in the study (Cohen's  $\kappa = 0.998$ ,  $p = 0.000$ ). In addition to counting how often each communication category was used, we also counted the total number of posts made (chat total) and the number of words used (chat length).

## 4.2. Output Variables

### 4.2.1. Team Performance

Ancona and Caldwell's (1992) definition of team performance is the extent to which a team can meet its output targets (e.g., quality, functionality, and reliability of outputs), the expectations of its members, or its cost and time goals (Ancona and Caldwell, 1992). For this study, the team performance metric consisted of the binary mapping of the task outcome (winning/losing). The team performance metric has been used as a dependent variable in our functional analysis of the collaboration to illustrate the role of the input factors (personality traits and communication patterns) and allow us to evaluate the constitution of those teams.

### 4.2.2. Perceived Collaboration Quality

To measure perceived collaboration quality, we use five metrics of team dynamics, which evaluated the participants' perceptions of their teams.

#### 4.2.2.1. Perceived Performance

The perceived performance metric addresses the question “*How well, in your opinion, did your team perform?*.” It was measured on a five-point Likert-scale from “*Very poorly*” (1) to “*Very well*” (5). The perceived performance variable defines the subjective layer of teamwork capability at the given task. The notion has been conceptualized as a multilevel process arising as the teammate engages in their individual and team-level task-work and teamwork processes (Kozlowski and Klein, 2000).

#### 4.2.2.2. Perceived Cohesion

The perceived cohesion metric addresses the question: “*How cohesive was your team?*,” measured using a similar 5-point Likert-scale. Perceived team cohesion, as a fringe term covering social relations, task relations, perceived unity, and emotions (Beal et al., 2003), contributes to our understanding of the emotional dimension of the teams, which is a rather subtle corollary facet of teamwork alongside other subjective measures. The study proposes that group members’ perceptions of their cohesion to a particular group are essential in the sense of belonging and feelings of morale (Bollen and Hoyle, 1990). More so, the meta-analysis by Beal et al. (2003) clarifying the construct relation between this particular subjective metric and team performance has denoted a high correlation between these factors across several studies on teams. This work has further established the importance of cohesion (including the subjective measurement) in team performance.

#### 4.2.2.3. Perceived Communication Quality

The perceived communication quality metric addresses the question: “*How well did your team communicate?*,” measured using a similar 5-point Likert-scale. Collecting the perception of the communication quality can help us encode important information about the participant’s beliefs toward how a team should function. It can also help disclose the way that the respective individuals engage in communication with the other team members and the way they perceive the communication ties (Cook et al., 2020). Differences in perception might uncover discrepancies between teammates’ viewpoints that can lead to the establishment of complex team interventions that intervene at multiple levels of the team formation and interaction processes (Wauben et al., 2011).

#### 4.2.2.4. Perceived Balance

The metric addresses the question: “*Did both members of your team contribute equally in your opinion?*” measured using a 3-point Likert-scale. The variable links with the staging of roles and responsibilities within a team, including how they distribute between teammates and the ways they get carried out against the team’s objectives (van de Water et al., 2008). To understand the relevance of the metric within the present study design, remember how entirely different the two roles are and how diametrically determinant they can contribute to teamwork. The top-down allocation of roles was, by itself, not a sufficient guarantee that the teammates’ behavior aligned with the given role. By assessing the aspect of perceived balance, through the lenses of the teammates, we could better understand what the

participants, and whether it was indeed a balanced act or whether a role was considered more demanding and accountable for the outcome than the other.

#### 4.2.2.5. Satisfaction

The metric addressed the question: “*Would you play with the same teammate again?*” measured using a 3-point Likert-scale. Satisfaction helps predict whether a combination of participants will more likely prefer to work with similar teammates in the future.

## 5. RESULTS

We divide our results into two themes: 1. performance, and 2. perceived collaboration quality.

### 1. Team performance:

- **Section 5.1** analyzes the effect of personality at team level<sup>13</sup>, comparing winning to losing teams to see if there may be a relationship between personality and performance. It reports the results of a Mann-Whitney U test and performs a regression to investigate the relationship between team traits and the likelihood of a team winning.
- **Section 5.2** analyzes the communication patterns using a one-way ANOVA to compare winners and losers, but also to compare the differences in behavior between the team roles.
- **Section 5.3** evaluates the impact on team performance of the participants’ socio-demographic characteristics, using Chi-square tests and regression analysis.

### 2. Perceived collaboration quality:

- **Section 5.4** assesses the relationship between personality traits and perception of collaboration quality, using correlation analysis for the individual traits.
- **Section 5.5** assesses the relationship between personality traits and perception of collaboration quality, using correlation analysis for the team traits.
- **Section 5.6** examines whether individual demographic characteristics played any role in people’s perception of their collaboration, using one-way ANOVAs.
- **Section 5.7** analyzes the relationship between the communication patterns and the collaboration quality metrics, also considering the roles of the Defuser and Lead Expert, using correlation analysis.

Given the many factors considered (e.g., considering 5 personality traits with 4 different aggregation metrics for team personality already results in 20 factors) and the many outcome measures, many statistical tests were performed. This may lead to Type I errors. Using Bonferroni corrections<sup>14</sup> to avoid Experiment wide Type I errors would reduce the power of

<sup>13</sup>Team, rather than individual level was used since it is usually the combination and interaction among individuals’ personalities that affects the team outcome, as evidenced by multiple studies [e.g., see Gilley et al.’s (2010) comprehensive review].

<sup>14</sup>Less conservative corrections such as Tukey are not possible due to the data often not meeting normality assumptions.

the statistical tests to such an extent that Type II errors would be highly likely and few insights would be gained<sup>15</sup>. We have therefore not applied such corrections (except in *post-hoc* pairwise comparisons). The study is exploratory in nature, and the statistical results presented provide initial insights that lead to hypotheses for follow-on studies.

### 5.1. Impact of Personality on Team Performance: Minimum Openness May Matter

Since there is no universally accepted way of aggregating team member personality traits into team personality traits, we used multiple, namely the average, minimum, maximum, and standard deviation. Each of these metrics was examined in isolation, as they are not independent. **Table 5** shows the mean (and standard deviation) of these four metrics for the winning and the losing teams. Minimum Openness was significantly better in winning teams (Mann-Whitney  $U = 485$ ,  $p = 0.02$ ). There were no other significant results<sup>16</sup>.

A binary logistic regression with the minimum metric<sup>17</sup> considered the effects of the teams personality on the likelihood of winning<sup>18</sup>. Given only 16 out of 60 teams won, the basic model only uses a constant with an accuracy of 73.3% (obtained by always predicting the team will lose). The logistic regression model was statistically significant,  $\chi^2_{(6)} = 13.60$ ,  $p = 0.034$ . The model explained 30% (Nagelkerke  $R^2$ ) of the variance in winning and correctly classified 77% of cases, including 38% of wins. Increasing minimum Openness and minimum Neuroticism were associated with an increased likelihood of winning [Openness:  $\text{Exp}(B) = 1.52$ ,  $\text{Wald} = 4.61$ ,  $p = 0.032$ ; Neuroticism:  $\text{Exp}(B) = 1.58$ ,  $\text{Wald} = 4.20$ ,  $p = 0.041$ ].

Our results indicate that in this kind of task (high-pressure, high-demand), minimum Openness to experience seems the most important factor among the Big-5 traits in helping the team to effectively manage the *ad-hoc* collaboration to find a winning solution within a limited time. This means that a crowdsourced, *ad-hoc*, and remote emergency response team will likely be more successful at executing a time-bounded novel task if both collaborators share high levels (minimum) of Openness to experience. The minimum level of this trait indicates that teams with individuals with low Openness are expected to hamper the collaboration regardless of whether the counterpart has very high levels of Openness and this is reasonably determined by the interdependence between roles.

<sup>15</sup>Additionally, as many measures were not independent, Bonferroni corrections would also have been less appropriate.

<sup>16</sup>Including no impact of Neuroticism or differences of standard deviation.

<sup>17</sup>We only performed the logistic regression with the minimum metric as minimum Openness was the only variable that was significant in the Mann-Whitney test, hence avoiding running multiple tests increasing the chances of Type I error.

<sup>18</sup>Hosmer and Lemeshow test was not significant, thus, the model assumptions were met.

### 5.2. Impact of Communication Patterns on Team Performance: Action and Response Help Teams Win

**Table 6** shows the number of posts per chat category for winners and losers, for winning and losing teams, and for Defusers and Lead Experts. As the role likely affects how participants communicate, we analyzed the communication pattern usage data at the individual level, with an output variable whether these people belonged to winning or losing teams. We analyzed the six chat categories (Uncertainty, Action, Response, Planning, Factual, Non-related), the chat length (in words) and the total number of chat posts between winners and losers using a one-way ANOVA. Winners used significantly more *Action* and *Response* statements [ $F_{action}(1,118) = 4.426$ ,  $p = 0.038$ ,  $F_{response}(1,118) = 4.983$ ,  $p = 0.027$ ].

A binary logistic regression model to predict whether a participant would win or lose was statistically significant [ $\chi^2_{(7)} = 14.86$ ,  $p = 0.038$ ]. The model explained 17% (Nagelkerke  $R^2$ ) of the variance in winning and correctly classified 78% of cases (25% wins). Increasing the *Action* and *Response* categories was associated with an increased likelihood of winning [ $\text{Exp}(B) = 1.28$ ,  $\text{Wald} = 5.35$ ,  $p = 0.021$ ;  $\text{Exp}(B) = 1.21$ ,  $\text{Wald} = 3.92$ ,  $p = 0.048$ , respectively]. Increasing the chat length was associated with a decreased likelihood of winning [ $\text{Exp}(B) = 0.97$ ,  $\text{Wald} = 4.04$ ,  $p = 0.044$ ]. These results seem to indicate that participants who gave feedback to one another and focused on discussing which action to take—rather than other types of communication—were able to finish the task and win the game. We also understand that the amount of chat is not a sufficient measure for success in online emergency response team settings since we could not find neither correlation nor causality between these variables.

Lead Experts used the *Action* category significantly more than Defusers [ $F_{action}(1,118) = 14.736$ ,  $p < 0.001$ ] whilst Defusers used the *Factual* category significantly more [ $F_{factual}(1,118) = 5.273$ ,  $p = 0.023$ ]. The Lead Experts are the ones with the map and would direct the Defusers to the appropriate path to defuse the bomb. Meanwhile, the Defusers may need to tell the Lead Experts where they are. There is a statistically significant difference in the chat categories, with Defusers on winning teams using a significantly higher proportion of *Factual* messages in their chat than those on losing teams (53 vs. 33%,  $p = 0.043$ ) and a lower proportion of *Uncertainty* messages (8 vs. 22%,  $p = 0.041$ ).

### 5.3. Impact of Socio-Demographic Characteristics on Performance

**Table 7** shows the demographics of winners vs. losers, excluding cases with very low frequency<sup>19</sup>. Pearson Chi-square tests show a significant association between gender and winning [ $\chi^2_{(1,N=119)} = 4.78$ ,  $p = 0.029$ ] and age and winning [ $\chi^2_{(3,N=120)} = 8.09$ ,  $p = 0.044$ ]. Men were more likely to win. A binary logistic regression model to predict whether a participant would win or lose based on gender was statistically significant [ $\chi^2_{(1)} = 5.12$ ,

<sup>19</sup>Namely prefer not say for gender, and British and Irish for nationality, all with frequency 1.

**TABLE 5 |** Mean (Stdev) of standard deviation, average, minimum, and maximum for personality traits for winning and losing teams.

		Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Winning Teams	StDev	1.06 (0.68)	1.41 (1.46)	1.15 (1.36)	1.50 (1.59)	1.10 (1.00)
	Average	8.13 (1.51)	7.75 (1.53)	5.75 (2.32)	6.94 (1.53)	4.22 (2.33)
	Min	7.38 (1.71)	6.75 (2.24)	4.94 (2.65)	5.88 (2.25)	3.44 (2.42)
	Max	8.87 (1.46)	8.75 (1.34)	6.56 (2.37)	8.00 (1.46)	5.00 (2.45)
Losing Teams	StDev	1.72 (1.36)	1.11 (1.32)	1.66 (1.52)	1.46 (1.25)	1.96 (1.88)
	Average	7.26 (1.60)	8.24 (1.41)	5.01 (1.55)	6.40 (1.26)	3.82 (1.74)
	Min	6.05 (2.22)	7.45 (1.95)	3.84 (1.80)	5.36 (1.79)	2.43 (1.37)
	Max	8.48 (1.42)	9.02 (1.39)	6.18 (1.97)	7.43 (1.25)	5.20 (2.78)

**TABLE 6 |** Mean (Stdev) of number of times chat categories were used by winners and losers, by winning and losing teams, by Defusers and Lead Experts, and total usage by each.

	Uncertainty	Action	Response	Planning	Factual	Non-related	Total
Winners	2.03 (3.10)	2.91 (4.85)	3.41 (3.77)	0.28 (0.58)	2.34 (2.89)	0.03 (0.18)	11.00 (11.15)
Losers	1.94 (2.30)	1.45 (2.60)	2.14 (2.29)	0.17 (0.49)	2.13 (2.49)	0.52 (2.82)	6.71 (11.00)
Winning teams	4.06 (4.71)	5.81 (6.66)	6.81 (7.08)	0.56 (1.09)	4.69 (4.47)	0.06 (0.25)	22.00 (20.41)
Losing teams	3.89 (3.27)	2.91 (3.67)	4.27 (4.01)	0.34 (0.77)	4.25 (4.21)	1.05 (4.08)	16.70 (11.55)
Defusers	1.62 (2.29)	0.72 (1.29)	2.32 (2.70)	0.27 (0.58)	2.72 (2.87)	0.07 (0.41)	7.70 (6.88)
Lead experts	2.32 (2.72)	2.97 (4.35)	2.63 (2.92)	0.13 (0.43)	1.65 (2.18)	0.72 (3.39)	10.42 (9.14)

$p = 0.024$ ]. However, it only explained 6% of the variance in winning and correctly classified 73.1% of cases only by always predicting losing. Being female was associated with a slightly decreased likelihood of winning [Exp(B) = -1.07, Wald = 4.53,  $p = 0.033$ ].

We also investigated whether adding gender to the model that uses personality to predict winning would improve the model. A binary logistic regression model to predict whether a participant would win or lose based on gender as well as team personality (in terms of minimum Openness and Neuroticism given the results from Section 5.1) was statistically significant [ $\chi^2_{(3)} = 27.97, p < 0.001$ ]. The model explained 31% (Nagelkerke  $R^2$ ) of the variance in winning and whilst correctly classifying 78.2% of cases. Being female was associated with a decreased likelihood of winning [Exp(B) = -1.31, Wald = 4.97,  $p = 0.026$ ]. Similar to our earlier results, increases in minimum Openness and Neuroticism were associated with an increased likelihood of winning [Exp(B) = 0.47, Wald = 11.92,  $p = 0.001$ ; Exp(B) = 0.52, Wald = 11.94,  $p = 0.001$ , respectively]. A similar model without Gender explained only 25% of the variance in winning, and reduced correct classification to 76.5%. Thus, gender mattered but less than personality. When age, nationality or education are added to the binary logistic model instead of gender, they are not significant.

### 5.4. Impact of Individuals Personality Traits on Perceived Collaboration Quality: Agreeableness May Be Helpful to Cope With Losing

Unfortunately, only 44 out of 120 participants (23 Lead Experts and 21 Defusers) completed the survey at the end of the

task, concerning their perception of their team’s Cohesion, Performance, Communication, Balance, and Satisfaction. All perceived collaboration metrics were positively correlated (see **Table 8**), overall and for winners. In contrast, for losers the correlations with Satisfaction were not significant (see **Table 8**), and Performance and Balance were also not correlated. So, losers may not always have attributed the bad performance to a poor balance in the team, nor always have been unwilling to keep working with a person even though the collaboration was not going well (according to the other metrics and the fact they lost).

Agreeableness significantly correlated with perceived Performance, Cohesion, and Balance. Neuroticism significantly correlated with only Balance (see **Table 9**). Considering only winners, there were no significant correlations between the personality traits and any metric. In contrast, losers had a significantly positive correlation on Agreeableness with Performance, Cohesion, and Communication. Furthermore, losers had a significant negative correlation on Conscientiousness with Communication. Agreeableness may have helped people to see their loss in a more positive light, making them feel more positively about their teams performance, communication and cohesion<sup>20,21</sup>. We do not know whether being more conscientious made losers feel worse about their teams communication, or whether

<sup>20</sup>This also means that Agreeableness needs to be considered when interpreting indirect measures of team collaboration quality as it may make them a less accurate reflection of actual collaboration.

<sup>21</sup>This seems more likely than that Agreeableness influenced the performance, communication, and cohesion itself, certainly given the lack of correlations for winners.

**TABLE 7 |** Demographics overall and of winners vs. losers (excluding prefer not to say for gender and nationality) and also for teams that include the same or different genders and nationalities.

	Gender				Nationality				Age				Education			
	Men	Women	Same	Differs	USA	India	Same	Differs	20–29	30–39	40–49	50+	HS	SC	Col	PG
N	78	41	33	27	67	51	33	27	23	56	26	15	9	9	87	15
Winners	33%	15%	30%	22%	19%	35%	27%	26%	22%	36%	27%	0%	11%	33%	28%	27%
Losers	67%	85%	70%	78%	81%	65%	73%	74%	78%	64%	73%	100%	89%	67%	72%	73%

**TABLE 8 |** Spearman correlations between perceived collaboration quality metrics, \*\* $p < 0.01$ , \* $p < 0.05$ .

		Performance	Cohesion	Communication	Balance	Satisfaction
All (N = 44)	Performance	1	0.751**	0.593**	0.449**	0.525**
	Cohesion	0.751**	1	0.649**	0.528**	0.502**
	Communication	0.593**	0.649**	1	0.506**	0.508**
	Balance	0.449**	0.528**	0.506**	1	0.389**
	Satisfaction	0.525**	0.502**	0.508**	0.398**	1
Winners (N = 24)	Performance	1	0.732**	0.648**	0.486*	0.568**
	Cohesion	0.732**	1	0.725**	0.512*	0.579**
	Communication	0.648**	0.725**	1	0.530**	0.646**
	Balance	0.486*	0.512*	0.530**	1	0.484*
	Satisfaction	0.568**	0.579**	0.646**	0.484*	1
Losers (N = 20)	Performance	1	0.734**	0.523*	0.302	0.299
	Cohesion	0.734**	1	0.514*	0.419	0.319
	Communication	0.523*	0.514*	1	0.470*	0.283
	Balance	0.302	0.419	0.470*	1	0.261
	Satisfaction	0.299	0.319	0.283	0.261	1

the team communication was influenced negatively by their Conscientiousness. The lack of a significant correlation for winners points toward the first explanation, with Conscientious people perhaps being more honest in assessing team communication quality.

### 5.5. Impact of the Teams Personality Traits on Perceived Collaboration Quality: The Positive Role of openness and Surprising Need for Conscientiousness Differences

We determined values for a teams perceived collaboration quality metrics by taking the average of its members, or only one member had provided their ratings by using that members ratings. Average and minimum Openness positively correlated with perceived performance<sup>22</sup> in line with earlier findings that Openness had a positive impact on the likelihood of a team winning. Maximum Agreeableness positively correlated with perceived performance<sup>23</sup>, in line with our earlier observations regarding the impact of Agreeableness on individuals opinions.

<sup>22</sup>Spearman correlations average Openness:  $r = 0.398$ ,  $p = 0.02$ ; minimum Openness  $r = 0.410$ ,  $p = 0.02$ .

<sup>23</sup>Spearman correlation:  $r = 0.400$ ,  $p = 0.02$ .

The most interesting result is the significant positive correlation of all perceived quality metrics with Conscientiousness standard deviation<sup>24,25</sup>.

A lower Conscientiousness standard deviation correlated with negative team’s feelings. In a dyad, the lowest Conscientiousness standard deviation is when two people work together who are very similar in Conscientiousness. For example, two highly conscientious people or two lowly conscientious people. Two lowly conscientious people working together may not result in a good collaboration. However, two highly conscientious people working together are likely to yield good performance. It seems that the best performance—from the team members’ point-of-view—for this particular type of task comes from two people differing in Conscientiousness working together.

### 5.6. Impact of Socio-Demographic Characteristics on Perceived Collaboration Quality: No Significant Result

Tables 10, 11 show the perceived collaboration quality metrics for the different genders, age groups, nationalities, and education

<sup>24</sup>Spearman correlations Performance:  $r = 0.644$ ,  $p < 0.0001$ ; Communication quality:  $r = 0.492$ ,  $p = 0.003$ ; Cohesion  $r = 0.403$ ,  $p = 0.02$ ; Balance:  $r = 0.448$ ,  $p = 0.008$ ; Satisfaction:  $r = 0.417$ ,  $p = 0.01$ .

<sup>25</sup>There was also a significant Spearman correlation for minimum Conscientiousness:  $r = -0.423$ ,  $p = 0.01$ .

**TABLE 9 |** Correlations between perceived collaboration quality metrics and personality traits, \*\* $p < 0.01$ , \* $p < 0.05$ .

		OPEN	CONS	EXTRO	AGR	NEUR
All (N = 44)	Performance	0.062	-0.187	0.044	0.434**	0.106
	Cohesion	0.050	-0.181	-0.088	0.319*	0.160
	Communication	-0.111	-0.256	-0.217	0.221	0.159
	Balance	-0.029	-0.203	-0.196	0.317*	0.318*
	Satisfaction	-0.003	-0.035	-0.074	0.032	-0.031
Winners (N = 24)	Performance	0.081	-0.099	0.064	0.289	-0.023
	Cohesion	0.053	-0.148	-0.006	0.241	0.013
	Communication	-0.068	-0.098	-0.239	-0.074	0.044
	Balance	-0.319	-0.302	-0.345	0.354	0.285
	Satisfaction	-0.086	0.144	-0.009	-0.072	-0.098
Losers (N = 20)	Performance	0.013	-0.336	0.017	0.761**	0.330
	Cohesion	0.021	-0.226	-0.162	0.456*	0.388
	Communication	-0.178	-0.551*	-0.159	0.547*	0.397
	Balance	0.315	-0.053	0.004	0.338	0.361
	Satisfaction	0.025	-0.233	-0.112	0.242	0.050

**TABLE 10 |** Mean (standard deviation) of collaboration quality metrics by gender and age, and also for teams that include the same or different genders.

Collaboration	Gender				Age			
	Men (32)	Women (12)	Same (20)	Differs (14)	20–29 (11)	30–39 (25)	40–49 (6)	50+ (2)
Performance	3.75 (1.27)	3.17 (1.53)	3.68 (1.17)	3.21 (1.53)	3.82 (0.87)	3.56 (1.50)	3.50 (1.64)	3.00 (1.41)
Cohesion	3.50 (1.19)	3.00 (1.28)	3.53 (1.09)	3.00 (1.32)	3.55 (1.04)	3.36 (1.22)	2.83 (1.72)	4.00 (0.00)
Communication	3.78 (1.24)	3.25 (1.29)	4.00 (1.06)	2.93 (1.27)	4.27 (0.65)	3.48 (1.33)	3.00 (1.67)	4.00 (0.00)
Balanced	1.03 (0.90)	1.08 (0.67)	1.10 (0.84)	0.89 (0.79)	1.09 (0.83)	1.12 (0.83)	0.33 (0.52)	2.00 (0.00)
Satisfied	1.38 (0.83)	1.08 (0.79)	1.23 (0.83)	1.32 (0.72)	1.27 (0.91)	1.20 (0.82)	1.83 (0.41)	1.00 (1.41)

levels. One-way ANOVAs showed no significant effect of socio-demographic variables on perceived team performance, cohesion, communication, balance, and satisfaction<sup>26</sup>. The averages on all metrics except for balance were a bit higher for men (which would make sense given the men had more often won), but this was not statistically significant, which is not surprising given the high variance and the sample size.

### 5.7. Impact of Communication Patterns on Perceived Collaboration Quality

We carried out a Spearman correlation test between the communication patterns (the number of occurrences of each communication category for the individual and their team) and the perceived collaboration quality (by individuals<sup>27</sup>).

*Satisfaction* was positively correlated with the *Factual* category ( $r = 0.308, p = 0.042$ , for both the individual and team), also for Defusers ( $r = 0.457, p = 0.037$ , for the individual), but not Lead Experts. So, members seemed more pleased when their team shared more facts, and Defusers particularly when they

shared more facts. Satisfaction was also positively correlated with *Planning* but only for Defusers ( $r = 0.437, p = 0.047$ , for the team). It suggests that Defusers were more pleased when the team planned toward the common goal (i.e., defusing the bomb on time).

*Performance* was positively correlated with the *Factual* category only for Defusers ( $r = 0.504, p = 0.020$ , for the team). The more cues were shared among the team members the better Defusers seemed to perceive the team performance.

*Balance* was negatively correlated with the *Uncertainty* category ( $r = -0.378, p = 0.011$ , for the individual), also for Lead Experts ( $r = -0.440, p = 0.036; r = -0.524, p = 0.010$ , for the individual and team respectively), but not for Defusers. The more questions the Lead Expert asked, and the more questions were asked in the team, the less balanced the Lead Experts seemed to perceive the collaboration.

Finally, *Communication* was positively correlated with the individual *Response* category for Defusers ( $r = 0.457, p = 0.028$ ), so the more responsive the Defuser was (e.g., in acknowledging actions they were going to perform), the better they regarded the team communication.

To summarize, several communication categories correlate with perceived collaboration quality, with the role in the team impacting which categories matter. For a good perceived

<sup>26</sup>There was a significant difference for education level on balance, but given the small numbers in all groups.

<sup>27</sup>Given the low number of teams where both members responded, we used the perceived collaboration quality at the individual level only.

**TABLE 11** | Mean (standard deviation) of collaboration quality metrics by nationality and education level, and also for teams that include the same or different nationalities.

Collab. Metrics	Nationality				Education Level			
	USA (16)	India (28)	Same (19)	Differs (15)	High Sch. (1)	Some Coll (3)	College (34)	Postgrad. (6)
Performance	3.19 (1.56)	3.82 (1.19)	3.76 (1.25)	3.13 (1.38)	3.00 (0.00)	4.00 (1.00)	3.62 (1.33)	3.33 (1.86)
Cohesion	3.31 (1.40)	3.39 (1.13)	3.53 (1.17)	3.03 (1.22)	3.00 (0.00)	3.33 (0.58)	3.44 (1.16)	3.00 (1.90)
Communication	3.31 (1.49)	3.82 (1.09)	3.68 (1.11)	3.40 (1.44)	2.00 (0.00)	4.67 (0.58)	3.71 (1.12)	3.00 (1.90)
Balanced	1.06 (0.93)	1.04 (0.79)	1.16 (0.78)	0.83 (0.84)	0.00 (0.00)	1.67 (0.58)	1.15 (0.78)	0.33 (0.82)
Satisfied	1.25 (0.86)	1.32 (0.82)	1.40 (0.76)	1.10 (0.81)	1.00 (0.00)	2.00 (0.00)	1.24 (0.82)	1.33 (1.03)

collaboration quality, it seemed important for Defusers to provide facts and neither the team nor the Lead Expert to ask too many questions.

### 5.8. Post-hoc Analysis on Impact of Culture

Given our participants mainly came from the USA and India, one may wonder whether there is an impact of culture. Firstly, whilst there is research to show that personality scales can be generalized across cultures (Rolland, 2002; Rammstedt and John, 2007), the distribution in cultures of personality traits differs. Sometimes therefore statine scores (Thorndike, 1982) are used for personality tests to normalize scores based on participants' country of origin. We did not do this, but did consider how the USA and India differ on personality scores, and whether this difference is visible in our participant sample. **Table 12** shows the personality scores for the USA and India from the literature, and the scores in our sample. In the literature, the main differences between these countries are on Extraversion and Agreeableness. In our sample, there were significant differences in Openness, Extraversion and Agreeableness between the sample from India and the USA<sup>28</sup>. If we had used stanine scoring normalizing based on the country averages from the literature, the difference between the scores in our sample would have been even bigger (given the averages for India where lower than those for the USA in the literature on these traits, and they already are higher than those for the USA in our sample). We conclude that crowd workers recruited through Mechanical Turk do not represent the average person from their countries. This is not surprising, as for example Burnham et al. (2018) found that Mechanical Turkers from the USA are lower in Extraversion than the general USA population (as was also the case in our sample). To be successful on Mechanical Turk, a certain level of conscientiousness is required (as many tasks require a certain success rate on previous tasks). Similarly, one could imagine that coming from India and working on an American platform requires a certain level of Openness to Experience.

There may also be an impact of whether people worked with somebody from their own culture in the task or another culture. We therefore considered whether there was a difference between same nationality teams and teams which differed in nationality on winning the task and on perceptions of collaboration quality (see descriptives in **Tables 7, 11**, respectively). There was

clearly no difference on winning or losing. The perception of collaboration quality seemed slightly better for same nationality teams (with higher means on all measures), but this difference was not statistically significant<sup>29</sup>.

## 6. DISCUSSION, LIMITATIONS, AND FUTURE WORK

### 6.1. Discussion

In this paper, we explored the impact of personality traits, demographics and communication patterns on a virtual collaborative task under time constraints for crowdsourced dyads. Our study observes how the crowd enacts pair-wise roles under pressure, adjusts its communication via chat, and shares common objectives while executing an artificial, video-game-inspired, cooperative time-bound task. Our goal is to use the knowledge from the observations gathered from the study as the basis for future work on AI-supported crowdsourcing of remote emergency response teams. The main results from our exploration, that will need to be verified in follow-on studies, are as follows:

- **Personality and team performance:** minimum Openness to experience seemed to affect the teams' ability to perform under time pressure. Comparatively, teams with higher minimum Openness levels performed better at the remote cooperative task.
- **Communication and team performance:** Communication patterns seemed to matter for team performance: better-performing crowd teams had more Action/Response statements than non-winning teams.
- **Demographics and team performance:** Gender seemed to influence performance, with men slightly more likely to win, however, gender influenced team performance less than the personality trait Openness to experience (minimum).
- **Personality and perception:** Crowd workers' Agreeableness and Conscientiousness likely shaped their perception of the collaboration. Furthermore, dyads that combined people differing in Conscientiousness were perceived by the participants themselves to perform better.

<sup>28</sup> Post-hoc test, Mann-Whitney  $U = 811.5, U = 611.0, U = 933.5$  respectively, with  $p < 0.001$  (and still significant if Bonferroni corrected).

<sup>29</sup> Perceived performance was significant at  $p < 0.05$ , but not when Bonferroni correction was applied.

**TABLE 12 |** Mean and standard deviation of the Big Five personality traits in the literature (Bartram, 2013) and in our sample data.

Data		Openness	Conscientiousness	Extraversion	Agreeableness	Emotional stability
Literature	USA	5.29 (2.05)	5.72 (2.03)	5.84 (2.09)	5.34 (1.97)	5.70 (2.05)
	USA	6.69 (2.19)	8.34 (1.95)	4.13 (2.02)	5.85 (1.83)	5.88 (2.92)
Our sample						
	India	8.55 (1.56)	7.80 (1.89)	6.71 (1.89)	7.43 (1.74)	6.35 (2.02)

- **Communication and perception:** Communication patterns also seemed to matter for perceived collaboration quality, with the role in the team impacting which categories mattered.

We weigh up these results and connect them with the broader teamwork literature in the coming sections.

### 6.1.1. Minimum Openness May Impact Teamwork in High-Stress Remote Tasks

Our study demonstrates that the trait of Openness to experience (specifically, its minimum level in a dyadic crowd team) may be a crucial feature for collaboration under pressure and time constraints. This result is novel to the field of team formation since several other studies (Thoms et al., 1996; Barrick et al., 1998; Coglisier et al., 2012; Curşeu et al., 2019) have found that other traits (Conscientiousness first, then Extraversion and Agreeableness) are the most relevant factors affecting team performance. There have been other studies on the effects of personality traits on team performance, such as by O’Neill and Allen (2011) indicating that the trait of Openness is negatively linked with performance *when the team is stable and long-term*, and when it has to perform large analytical tasks such as software engineering. In view of O’Neill and Allen’s (2011) study, we read our results as being strongly conditioned by the chosen task type. By highlighting the importance of the trait of Openness, our study helps shed light on the differences that distinguish online *ad-hoc* teams for high-pressure, high-stake tasks, from classical team settings.

Adaptation, as a collateral personality feature of individuals with high Openness to experience, is indeed considered useful in teamwork (Gallivan, 2004), especially in situations of high stress, high-stake and limited time. Moreover, intellectual curiosity with regards to new circumstances is a characteristic observed in people with high Openness to experience (McCrae, 1987); this same trait is closely related to team creativity (Schilpzand et al., 2010). Substantiated by literature (McCrae, 1987; Schilpzand et al., 2010), our results suggest that Openness may act as a more influential factor than task familiarity in determining the success of the team.

### 6.1.2. Focused Communication Patterns Get the Teams Going

From the results of the analysis of the collaboration, patterns emerge that people who completed the challenge had substantially more Action/Response statements in their chat logs. Thus, they were more effective at communicating with their teammate and promptly came up with clear instructions that helped solve the task on time. Successful participants under pressure used the chat to find a solution right away. Furthermore, winning Defuser predominantly used factual

statements. Winning Defusers paid attention to the directives given by their paired teammates (Lead Experts) and responded over the chat by describing where they were at that point in the maze. These results seem to indicate the importance of *focused communication* (with the focus being on efficiency and action clarity), especially when the stakes are high and time-bound. The identification of collaboration patterns has also uncovered tangible clues on how winning individuals intervene during the novel, high-pressure circumstances. Even though communication styles were not communicated explicitly at the start of the task, some participants were more apt at adopting suitable conversational styles as they cooperated and learned from the activity. These findings corroborate other (quasi) longitudinal observations of the long-term impact of risk communication and emergency response measures (Heath and Palenchar, 2000) indicating that citizens are willing to become knowledgeable of emergency response measures and proactively contribute to community relations.

### 6.1.3. Agreeableness and Conscientiousness Likely Shape the Perception of Collaboration

In our study, highly agreeable people seem to deal better with losing, reflecting more positively on perceived performance, cohesion, and communication. Agreeableness has a social orientation (Bradley et al., 2013) and the trait faceted with trust, altruism, and humility (Matsumoto and Juang, 2016). As highly agreeable people tend to be more sympathetic toward others (Thompson, 2008) and more humble, this may have made them more forgiving toward their teammates and themselves on these aspects. We also found that individuals in teams heterogeneous on Conscientiousness felt better toward the collaboration. Hence, Conscientiousness, at least for high-pressure tasks, is better distributed across teams to improve the perception of teamwork. Making such teams that are heterogeneous in Conscientiousness does not have to be detrimental to actual performance, as shown by our other results as well as Mohammed and Angell (2003). Our result conflicts with that of Gevers and Peeters (2009) who showed that diverse levels of Conscientiousness were negatively linked with teammates’ satisfaction. It may be due to the nature of the task since homogeneous high Conscientiousness might have led both the Defuser and the Lead Expert to be overly cautious; however, further studies should investigate the extent of our findings.

### 6.1.4. Communication Patterns Aligned With Team Roles Matter for the Perception of Collaboration

Communication patterns seemed to matter for the perceived collaboration quality, but this depended heavily on team role. Defusers seemed more satisfied with the collaboration when both

themselves and the team used more Factual statements, Lead Experts seemed less satisfied when using Uncertainty statements. These results indicate the importance of team roles and how they are enacted and perceived by teammates. In this instance, the two team roles had distinct and interdependent duties. These reflected the communication patterns that the participants used and preferred (or disliked) above all. In the presence of such distinct team roles, the participants seem to have expected certain communication patterns from their teammates, and these greatly depended on what part of the information they had access to. Defining clear roles is important, as team role clarity improves collaboration (Aritzeta et al., 2005) and communication styles aligned with team roles matter for effective and satisfactory teamwork [as shown in this paper, and in line with (De Vries et al., 2006)]. It may be even more vital in high-pressure tasks with high interdependence.

### 6.1.5. Gender May Impact Collaboration Though Less Than Personality

Gender seemed to impact team performance, with men slightly more likely to win than women. We considered whether there may have been personality differences. We did not find a statistically significant difference in overall personality traits between genders in this sample. There is some evidence in the literature that there may be a difference in sub-facets of Openness (Weisberg et al., 2011). We also considered whether this is a side effect of the different proportions of men in the sample. More men would result in more teams with men being homogeneous in gender. However, we did not find a significant difference in performance between homogeneous and heterogeneous genders (see **Table 7** for descriptives for same gender teams and teams with different genders). Apestequia et al. (2012) considered the impact of gender on teamwork in an investment game setting. They argued that a decreased performance in homogeneous female teams is explained by differences in decision making, with women being less aggressive and more focused on social sustainability.

We also considered whether gender homogeneity impacted perceptions of collaboration quality (see **Table 10** for descriptives). There was a significant impact only on Communication (*post-hoc*, Mann Whitney  $U = 268$ ,  $p < 0.005$ , Bonferroni corrected), with Communication being appreciated more in same gender teams. As there is a big difference between India and the USA in gender equality (USA is 30<sup>th</sup> (out of 156) in the Global Gender Gap Index (Sharma et al., 2021) compared to India only being 140<sup>th</sup>), we also considered the impact of gender homogeneity when teams were diverse in nationality. For teams diverse in gender, there was a significant impact of nationality homogeneity on Cohesion and Balance (*post-hoc*, Mann Whitney  $U = 28$ ,  $p < 0.05$ , Bonferroni corrected) and similar trends for Communication and Performance ( $p = 0.1$  after Bonferroni correction), with all being perceived better for same nationality teams. We considered whether the impact of gender on winning we found may be partially due to women being more likely to have been in diverse gender teams, and collaboration issues having occurred in such teams when the teams were mixed in nationality. However, this was not supported by the data. Further

studies are needed to investigate possible cultural factors and their interaction with gender homogeneity. However, given the impact gender may have, gender diversity in teams should be encouraged (Díaz-García et al., 2013).

## 6.2. Limitations

### 6.2.1. Exploratory Study

As explained above, the study performed was exploratory in nature. Follow-on studies are needed to confirm the results found. The findings from our study can provide the hypotheses for such studies.

### 6.2.2. Matchmaking System

One of the primary limitations of this study comes from the matchmaking part of the system. We paired participants following a simple first-in-first-out queuing fashion and did not consider user features. This study design choice matched the micro-tasking nature of crowdsourcing and its asynchronous environment, characteristics typical to platforms like Amazon Mechanical Turk. Random matching proved to be an effective solution to the problem of pairing virtual users into *ad-hoc* teams fast and based on availability, and for this reason easily applicable in emergencies. However, this matching limited the control over team formation, rendering the present study observational. For future studies, we plan to test other types of matchmaking criteria. For example, using heuristic algorithms similar to Irvin's Stable Roommate Problem (Irving, 1985) that would assist the matchmaking process according to pre-defined criteria. Other matching systems, such as AI (machine learning and features extraction), could also be used as baselines.

### 6.2.3. Metrics and Sample

Another limitation of this study is the one associated with the dataset generated from the user outputs and their willingness to give away credible information on their personality traits, demographic data, and experience in the game. We plan to strengthen this area of the research by implementing additional types of secondary data collection systems, such as behavioral, contextual, and sensor data, to help validate and enrich the information gathered about the participants. Different user groups (e.g., students, remote developers, and incident response volunteers) should partake in future studies.

Additionally, our study design did not implement exclusion criteria such as required English proficiency levels nor relied upon pre-screening to filter crowd workers on the basis of their reputation and/or a number of successful HITs. Varying levels of English may have impacted the results. However, most participants reported having completed a College education and the education language at College in all participants' countries (USA, India, UK, Ireland) is English, so we have some confidence that the English level was sufficient not to inhibit communication. We also did not notice clear communication issues due to language in the chats. Nevertheless, future studies will include a test to ensure an appropriate English proficiency level. The absence of pre-screening on English also has a positive aspect, as means our study can be generalized to emergency crises where

English is not necessarily the native language whilst still being used for basic virtual communication via chat.

Finally, our sample consisted of predominately male, American, and Indian AMT workers. The sample used for the results likely impacted participants' collaboration and performance. Although we accounted for some of these socio-demographic characteristics (of which gender was significant), we acknowledge the limitations of the dataset derived from the AMT sample. Other types of remote crowd workers from other platforms should experiment with the tool to test for the generalisability of the findings to other portions of the population.

#### 6.2.4. Task, Timer, and Features

The results gathered from the experiments on a single task provide a limited range of conclusions and levels of abstraction to other domains unless other high-stress scenarios could be tested and compared. We plan to implement several types of high-stress tasks. For instance, real-time translation or visual puzzle games would generate more diverse data. They would also quantify the extent to which the choice of task design impacts team collaboration. Another limitation is the lack of manipulation checks for the perceived realism and urgency of the task. It is possible that those workers who did not approach the task seriously might have behaved differently in situations of authentic danger and gravity. Future work should apply similar methodologies and observations to real-life remote emergency situations to be able to test the generalizability of our findings<sup>30</sup>. As part of the development stage, we ran several pilot studies to improve the initial task design and make the instructions clear and understandable for the participating crowd workers.

In the process, we omitted multiple elements present in the original version of the module. We tested different countdowns during the pre-study phase with real users. We settled for a timelimit of 400s as it allowed participants to familiarize themselves with the task interface, chat with one another, and execute the task. Time limits can still be the subject of further testing to evaluate the user's reaction times. We deliberately excluded some of the original elements of the maze module from the video game (i.e., the count of strikes or penalty points for hitting the invisible blocks when crossing the walls, the view of the multiple mazes from the Lead Expert manual, etc.). Tweaking in-game parameters will help uncovering differences in behavior and collaboration that we could not identify by running a single study design. In our experiments, the maze's walls were made invisible to the Defuser while still detectable through object collision. In future studies, and as part of the task improvements, we aim to bring back some of the original features and to assess their significance.

### 6.3. Implications and Future Work

#### 6.3.1. AI Support for Team Formation in Emergency Response

There has been growing research on AI supported team formation, where AI programs allocate workers or learners

to teams (Lykourantzou et al., 2013; Odo et al., 2019a). Clearly, the task impacts what team attributes matter for good actual and perceived performance and collaboration. For the emergency task studied in this paper, our primary finding concerns the importance of the trait of Openness to Experience (minimum). When developing an AI group formation system, this can be incorporated (e.g., in the criteria used for automated team formation), ensuring emergency response team have high minimum Openness to Experience, and diverting crowd workers with low Openness to more suitable tasks. Pre-screening and selection procedures are not new to disaster management, but our findings indicate that certain personality traits affect emergency teamwork, and this goes beyond the more common filtering criteria used such as reputation and trust (Javaid et al., 2013). More so, previous research on the effects of personality traits in teamwork did not consider the impact of the task type under stress (Thoms et al., 1996; Barrick et al., 1998; Cogliser et al., 2012; Curşeu et al., 2019), particularly in cases of emergency response. The sample of crowd workers used in this study helped us understand how pairs of non-familiar and dispersed users act together when presented with an unseen challenge. By utilizing AI to infer the crowd's attributes through their interactions, intelligent systems can learn to adjust to their needs and capabilities in times of emergency and suggest collaborators for a better fit.

The results from this specific approach are beneficial to the crowdsourcing and online work fields that are becoming ever so relevant due to recent and significant changes in the way we live and work. In the Ukrainian conflict of 2022, volunteers of remote rescue operations based in the USA allocated buses to civilians making requests for help online and helping save countless lives (Mark et al., 2022). By remote communication and real-life GPS updates, citizens from far away aided the evacuation of many citizens by identifying grounds hit by shelling and bombing. Following tragic examples like this one, researchers and industry can weigh the power of AI to aid the team formation process of remote emergency crowd teams and assist with organizing rescue units during high-stress, life-threatening situations.

#### 6.3.2. Conversational AI Support for Remote Emergency Response Teams

The analysis of the communication patterns clearly indicated that not all teams focused on the task execution correctly since some adopted less-than-optimal communication strategies. Our results provide insights into which communication acts may be important which can be used by an AI system to monitor and moderate remote collaboration and intervene when needed. With the implementation of machine learning models, future crowdsourcing tools specialized in emergency response can augment the chat functionality by deploying conversational AI (Battineni et al., 2020) (as an example) moderating users' communication patterns. With the stark improvements in Natural Language Generation, Understanding, and Processing, and the increasingly reduced costs of production thanks to open-source software community (Adamopoulou and Moussiades, 2020), most forms of crowdsourced self-organized teams (e.g., neighborhood watch Bakker et al., 2012) could themselves

<sup>30</sup>However, there are clear ethical issues with this.

incorporate, maintain, and improve machine learning models for emergency response conversational AI initially trained on annotations and knowledge such as the one we present.

We note that personality traits seemed to affect the perception of the collaboration. Although system evaluations usually pursue metrics similar to ours (e.g., effectiveness, efficiency, and reliability), team performance is only part of the equation. While a team can successfully reach a goal on time, the perception of teamwork is not always directly proportional to that outcome. What individuals think, interpret, and how they respond to changes can be conditioned by personality factors. In this study, we observe the interaction between personality and communication patterns. With defined team roles and interdependency, people with certain personality traits are likely to expect from others certain communication styles. Further, personality seems to have determined the propensity for more or less rigor and clarity in the communication. Considering the numerous variables at play and the increased reliance on crowdsourcing for rescue operations and emergency response (Marc Cieslak, 2022), we advocate for the development of adaptive and personalized intelligent systems. AI-aided emergency response can provide support and knowledge to teams according to the individual and group needs to alleviate stress and improve community participation. Emotional support could be tailored to the individuals and made accessible and private in critical emergency settings addressing the lack of sensemaking and trust emerging from periods of stress, trauma, and danger.

## 7. CONCLUSION

In this study, 60 crowd dyads collaborated in a high-pressure, computer-mediated task. The study required them to play complementary roles in a time-bounded critical scenario. We explored the possible impact of the participants' personality, socio-demographic factors, and communication patterns on team performance and perceived collaboration quality. Results from our exploratory study suggest that teams scoring high on the personality trait of Openness (meaning that the minimum Openness of winning teams was higher than in the losing teams) performed overall better in the execution of this

high-pressure task. The analysis of the team communication patterns suggest that teams communicating more through action-response loops were more likely to win the game. Different levels of Agreeableness and Conscientiousness likely shaped the perception of collaboration with highly agreeable people coping better with losing. Teams heterogeneous on Conscientiousness seemed to feel better about the teamwork. Communication patterns seemed to matter for the perceived collaboration quality, but this was highly role-dependent, showing that communication styles aligned with team roles matter for effective and satisfactory teamwork. We can learn from these exploratory results that the perception of the collaboration may differ depending on personality traits and the communication patterns shared among remote teammates. So, intelligent crowdsourcing-aided emergency response technology may need to consider individuals' viewpoints and provide adequate support for the crowd needs. Our findings support future research on computer-based collaboration under pressure. It shows ways to tailor the development of AI as accessible support in crowdsourcing emergency response aiding with team formation, conversational support, and adaptation. Future work will confirm the findings and evaluate other types of high-stress tasks, time limits, and parameters for team formation to advance the findings presented here.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## REFERENCES

- Abdul-Razik, M. S., Kaity, A. M., Banafaa, N. S., and El-Hady, G. W. (2021). Disaster response in a civil war: lessons on local hospitals capacity. the case of yemen. *Int. J. Healthcare Manag.* 14, 99–106. doi: 10.1080/20479700.2019.1616386
- Abu-Elkheir, M., Hassanein, H. S., and Oteafy, S. M. (2016). "Enhancing emergency response systems through leveraging crowdsensing and heterogeneous data," in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)* (Paphos: IEEE), 188–193.
- Adamopoulou, E., and Moussiades, L. (2020). "An overview of chatbot technology," in *IFIP International Conference on Artificial Intelligence Applications and Innovations* (Crete: Springer), 373–383.
- Akman, I., Misra, S., and Altindag, T. (2011). The impact of cognitive and socio-demographic factors at meetings during software development process. *Tehnički vjesnik* 18, 51–56. Available online at: <https://hrcak.srce.hr/65924>
- Alexander, D. (2003). Towards the development of standards in emergency management training and education. *Disaster Prevent. Manag.* 12, 113–123. doi: 10.1108/09653560310474223
- Allen, D. W. (2011). *Getting to Know ArcGIS: Modelbuilder, Vol. 380*. Seattle, WA: Esri Press Redlands.
- Ancona, D. G., and Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organ. Sci.* 3, 321–341. doi: 10.1287/orsc.3.3.321
- Anderson, M. (2020). *How the Help Desk Can Support the Security Team* (Ph.D. thesis), Utica College.

- Apestequia, J., Azmat, G., and Iriberry, N. (2012). The impact of gender composition on team performance and decision making: evidence from the field. *Manag. Sci.* 58, 78–93. doi: 10.1287/mnsc.1110.1348
- Archibold, R. C. (2003). *A Nation at war: The New Organizations; Papers Debate Use of Teams to React Quickly to Terror*. New York: Time.
- Aritzeta, A., Ayestaran, S., and Swailes, S. (2005). Team role preference and conflict management styles. *Int. J. Conflict Manag.* 16, 157–182. doi: 10.1108/eb022927
- Bakker, J., Denters, B., Oude Vrielink, M., and Klok, P.-J. (2012). Citizens initiatives: How local governments fill their facilitative role. *Local Govern. Stud.* 38, 395–414. doi: 10.1080/03003930.2012.698240
- Baldwin, S., and Woods, P. A. (1994). Case management and needs assessment: some issues of concern for the caring professions. *J. Mental Health* 3, 311–322. doi: 10.3109/09638239408997941
- Barrick, M. R., Stewart, G. L., Neubert, M. J., and Mount, M. K. (1998). Relating member ability and personality to work-team processes and team *Effectiveness* 83, 377–391. doi: 10.1037/0021-9010.83.3.377
- Bartram, D. (2013). Scalar equivalence of opq32: Big five profiles of 31 countries. *J. Cross Cult. Psychol.* 44, 61–83. doi: 10.1177/0022022111430258
- Battinini, G., Chintalapudi, N., and Amenta, F. (2020). Ai chatbot design during an epidemic like the novel coronavirus. *Healthcare* 8, 154. doi: 10.3390/healthcare8020154
- Beal, D. J., Cohen, R. R., Burke, M. J., and McLendon, C. L. (2003). Cohesion and performance in groups: a meta-analytic clarification of construct relations. *J. Appl. Psychol.* 88, 989. doi: 10.1037/0021-9010.88.6.989
- Belbin, R. M. (2012). *Team Roles at Work*. Oxon: Routledge.
- Bell, S. T., Fisher, D. M., Brown, S. G., and Mann, K. E. (2018). An approach for conducting actionable research with extreme teams. *J. Manag.* 44, 2740–2765. doi: 10.1177/0149206316653805
- Bergen, M. (2020). *Google outage reignites worries about smart home without backups*. Available online at: <https://www.bloomberg.com/news/newsletters/2020-12-16/google-outage-reignites-worries-about-smart-home-without-backups>
- Bjerge, B., Clark, N., Fisker, P., and Raju, E. (2016). Technology and information sharing in disaster relief. *PLoS ONE* 11, e0161783. doi: 10.1371/journal.pone.0161783
- Bollen, K. A., and Hoyle, R. H. (1990). Perceived cohesion: a conceptual and empirical examination. *Soc. Forces* 69, 479–504. doi: 10.2307/2579670
- Borman, W. C., Rosse, R. L., and Abrahams, N. M. (1980). An empirical construct validity approach to studying predictor-job performance links. *J. Appl. Psychol.* 65, 662. doi: 10.1037/0021-9010.65.6.662
- Bowers, C. A., Jentsch, F., Salas, E., and Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Hum. Factors* 40, 672–679. doi: 10.1518/001872098779649265
- Boyd, R., and Brown, T. (2005). Pilot study of myers briggs type indicator personality profiling in emergency department senior medical staff. *Emergency Med. Aust.* 17, 200–203. doi: 10.1111/j.1742-6723.2005.00723.x
- Boyle, G. J. (1995). Myers-briggs type indicator (mbti): some psychometric limitations. *Aust. Psychol.* 30, 71–74. doi: 10.1111/j.1742-9544.1995.tb01750.x
- Bradley, B. H., Baur, J. E., Banford, C. G., and Postlethwaite, B. E. (2013). Team players and collective performance: how agreeableness affects team performance over time. *Small Group Res.* 44, 680–711. doi: 10.1177/1046496413507609
- Brennan, M. A., and Flint, C. G. (2007). Uncovering the hidden dimensions of rural disaster mitigation: capacity building through community emergency response teams. *J. Rural Soc. Sci.* 22, 7. Available online at: <https://egrove.olemiss.edu/jrss/vol22/iss2/7>
- Briggs, S. M. (2005). Disaster management teams. *Curr. Opin. Crit. Care* 11, 585–589. doi: 10.1097/01.ccx.00001186916.92757.ab
- Burnham, M. J., Le, Y. K., and Piedmont, R. L. (2018). Who is mturk? personal characteristics and sample consistency of these online workers. *Mental Health Religion Cult.* 21, 934–944. doi: 10.1080/13674676.2018.1486394
- Capraro, R. M., and Capraro, M. M. (2002). Myers-briggs type indicator score reliability across: studies a meta-analytic reliability generalization study. *Educ. Psychol. Meas.* 62, 590–602. doi: 10.1177/0013164402062004004
- Careem, M., De Silva, C., De Silva, R., Raschid, L., and Weerawarana, S. (2006). “Sahana: overview of a disaster management system,” in *2006 International Conference on Information and Automation* (Colombo: IEEE), 361–366.
- Carver, L., and Turoff, M. (2007). Human-computer interaction: the human and computer as a team in emergency management information systems. *Commun. ACM* 50, 33–38. doi: 10.1145/1226736.1226761
- Chau, M. M. (2020). Rapid response to a tree seed conservation challenge in hawaii ‘i through crowdsourcing, citizen science, and community engagement. *J. Sustain. Forestry* 1–19. doi: 10.1080/10549811.2020.1791186. [Epub ahead of print].
- Chen, L., and Miller-Hooks, E. (2012). Optimal team deployment in urban search and rescue. *Transport. Res. B Methodol.* 46, 984–999. doi: 10.1016/j.trb.2012.03.004
- Chen, R., Sharman, R., Rao, H. R., and Upadhyaya, S. J. (2008). Coordination in emergency response management. *Commun. ACM* 51, 66–73. doi: 10.1145/1342327.1342340
- Coglisier, C. C., Gardner, W. L., Gavin, M. B., and Broberg, J. C. (2012). Big five personality factors and leader emergence in virtual teams: relationships with team trustworthiness, member performance contributions, and team performance. *Group Organ. Manag.* 37, 752–784. doi: 10.1177/1059601112464266
- Colovic, A., Caloffi, A., and Rossi, F. (2022). Crowdsourcing and covid-19: how public administrations mobilize crowds to find solutions to problems posed by the pandemic. *Public Adm Rev.* doi: 10.1111/puar.13489. [Epub ahead of print].
- Cook, A., Zill, A., and Meyer, B. (2020). Perceiving leadership structures in teams: Effects of cognitive schemas and perceived communication. *Small Group Res.* 25, 251–287. doi: 10.1177/1046496420950480
- Curşeu, P. L., Ilies, R., Virgă, D., Maricuţoiu, L., and Sava, F. A. (2019). Personality characteristics that are valued in teams: not always “more is better”? *Int. J. Psychol.* 54, 638–649. doi: 10.1002/ijop.12511
- Davaslioglu, K., Pokorny, B., Sagduyu, Y. E., Molintas, H., Soltani, S., Grossman, R., et al. (2019). “Measuring the collective allostatic load,” in *2019 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* (Las Vegas, NV: IEEE), 93–99.
- De Vries, R. E., Van den Hooff, B., and de Ridder, J. A. (2006). Explaining knowledge sharing: the role of team communication styles, job satisfaction, and performance beliefs. *Commun. Res.* 33, 115–135. doi: 10.1177/0093650205285366
- De Wit, F. R., Greer, L. L., and Jehn, K. A. (2012). The paradox of intragroup conflict: a meta-analysis. *J. Appl. Psychol.* 97, 360. doi: 10.1037/a0024844
- Díaz-García, C., González-Moreno, A., and Jose Saez-Martinez, F. (2013). Gender diversity within r&d teams: its impact on radicalness of innovation. *Innovation* 15, 149–160. doi: 10.5172/impp.2013.15.2.149
- Difallah, D., Filatova, E., and Ipeirotis, P. (2018). “Demographics and dynamics of mechanical turk workers,” in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. (Marina Del Rey, CA: ACM), 135–143.
- Dilmaghani, R. B., and Rao, R. R. (2006). “On designing communication networks for emergency situations,” in *2006 IEEE International Symposium on Technology and Society* (Queens, NY: IEEE), 1–8.
- Driskell, T., Burke, S., Driskell, J., Salas, E., and Neuberger, L. (2014). Stealing the team: assessing individual and team functioning “at a distance.” *Military Psychol.* 29, 12–18.
- Driskell, T., Salas, E., and Driskell, J. E. (2018). Teams in extreme environments: Alterations in team development and teamwork. *Hum. Resour. Manag. Rev.* 28, 434–449. doi: 10.1016/j.hrmmr.2017.01.002
- Ducao, A. A. B. (2013). *OpenIR [Open Infrared]: enhancing environmental monitoring through accessible remote sensing, in Indonesia and beyond* (Ph.D. thesis), Massachusetts Institute of Technology.
- Ellis, A. P., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., and Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: effects on cognitive and skill-based outcomes. *Pers. Psychol.* 58, 641–672. doi: 10.1111/j.1744-6570.2005.00617.x
- Elsafoury, F. (2020). “Teargas, water cannons and twitter: a case study on detecting protest repression events in turkey 2013,” in *Text2Story@ ECIR*. (Lisbon), 5–13.

- Ernst, C., Mladenow, A., and Strauss, C. (2017). Collaboration and crowdsourcing in emergency management. *Int. J. Pervasive Comput. Commun.* 13, 176–193. doi: 10.1108/IJPC-03-2017-0026
- Etheridge, J. C., Moyal-Smith, R., Sonnay, Y., Brindle, M. E., Yong, T. T., Tan, H. K., et al. (2022). Non-technical skills in surgery during the covid-19 pandemic: an observational study. *Int. J. Surg.* 98, 106210. doi: 10.1016/j.ijss.2021.106210
- Farkas, J., and Neumayer, C. (2017). 'stop fake hate profiles on facebook: Challenges for crowdsourced activism on social media. *First Monday.* 22. doi: 10.5210/fm.v22i9.8042
- Flin, R., and Slaven, G. (1996). Personality and emergency command ability. *Disaster Prevent. Manag.* 5, 40–46. doi: 10.1108/09653569610109550
- Floch, J., Angermann, M., Jennings, E., and Roddy, M. (2012). "Exploring cooperating smart spaces for efficient collaboration in disaster management," in *ISCRAM*. (Vancouver, BC).
- Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *Am. Psychol.* 39, 885. doi: 10.1037/0003-066X.39.8.885
- Friede, A. M. (2022). In defence of the baltic sea region:(non-) allied policy responses to the exogenous shock of the ukrainian crisis. *Eur. Security* 1–23. doi: 10.1080/09662839.2022.2031990. [Epub ahead of print].
- Galbraith, J. R., and Lawler, E. E. (1993). *Organizing for the Future: The New Logic for Managing Complex Organizations*. Durham: Jossey-Bass Inc. Publisher.
- Gallivan, M. J. (2004). Examining it professionals' adaptation to technological change: The influence of gender and personal attributes. *ACM SIGMIS Database* 35, 28–49. doi: 10.1145/1017114.1017119
- Gardner, H. K. (2012). Performance pressure as a double-edged sword: enhancing team motivation but undermining the use of team knowledge. *Administrat. Sci. Q.* 57, 1–46. doi: 10.1177/0001839212446454
- Gay-Antaki, M. (2021). Stories from the IPCC: an essay on climate science in fourteen questions. *Glob. Environ. Change* 71, 102384. doi: 10.1016/j.gloenvcha.2021.102384
- Gersick, C. J., and Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organ. Behav. Hum. Decis. Process.* 47, 65–97. doi: 10.1016/0749-5978(90)90047-D
- Gevers, J. M. P., and Peeters, M. A. G. (2009). A pleasure working together? the effects of dissimilarity in team member conscientiousness on team temporal processes and individual satisfaction. *J. Organ. Behav.* 30, 379–400. doi: 10.1002/job.544
- Gilley, J. W., Morris, M. L., Waite, A. M., Coates, T., and Veliquette, A. (2010). Integrated theoretical model for building effective teams. *Adv. Dev. Hum. Resour.* 12, 7–28. doi: 10.1177/1523422310365309
- Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *J. Pers. Soc. Psychol.* 59, 1216. doi: 10.1037/0022-3514.59.6.1216
- Hackman, J. R. (1990). *Groups that work and those that don't*. Number E10 H123. Jossey-Bass.
- Hackman, J. R., and Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. *Adv. Exp. Soc. Psychol.* 8, 45–99. doi: 10.1016/S0065-2601(08)60248-8
- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. (2018). "A data-driven analysis of workers' earnings on amazon mechanical turk," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. (Montreal, QC), 1–14.
- Harrison, A. A., and Connors, M. M. (1984). Groups in exotic environments. *Adv. Exp. Soc. Psychol.* 18, 49–87. doi: 10.1016/S0065-2601(08)60142-2
- Hart, J. D., Piumsomboon, T., Lawrence, L., Lee, G. A., Smith, R. T., and Billinghurst, M. (2018). "Demonstrating emotion sharing and augmentation in cooperative virtual reality games," in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. (Melbourne, VIC), 117–120.
- Heath, C., and Luff, P. (1992). Collaboration and control/crisis management and multimedia technology in london underground line control rooms. *Comput. Support. Cooperative Work* 1, 69–94. doi: 10.1007/BF00752451
- Heath, R. L., and Palenchar, M. (2000). Community relations and risk communication: a longitudinal study of the impact of emergency response messages. *J. Public Relat. Res.* 12, 131–161. doi: 10.1207/S1532754XJPRR1202\_1
- Heinzlman, J., and Waters, C. (2010). *Crowdsourcing Crisis Information in Disaster-Affected Haiti*. Washington, DC: U.S. Institute of Peace.
- Helmreich, R. L. (1967). *Prolonged stress in sealab ii: A field study of individual and group reactions*. Technical report, Yale University New Haven Ct.
- Hemphill, L., and Roback, A. J. (2014). "Tweet acts: how constituents lobby congress via twitter," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*. (Baltimore, MD: ACM), 1200–1210.
- Herstein, J. J., Schwedhelm, M. M., Vasa, A., Biddinger, P. D., and Hewlett, A. L. (2021). Emergency preparedness: What is the future? *Antimicrob. Stewardship Healthcare Epidemiol.* 1, e29. doi: 10.1017/ash.2021.190
- Highhouse, S., Wang, Y., and Zhang, D. C. (2022). Is risk propensity unique from the big five factors of personality? a meta-analytic investigation. *J. Res. Pers.* 98, 104206. doi: 10.1016/j.jrp.2022.104206
- Hossain, A., Mirza, F., Naeem, M. A., and Gutierrez, J. (2017). "A crowd sourced framework for neighbour assisted medical emergency system," in *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)* (Melbourne, VIC: IEEE), 1–6.
- Hughes, J. A., Randall, D., and Shapiro, D. (1992). "Faltering from ethnography to design," in *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*. (Toronto, ON: ACM), 115–122.
- Ikizer, G., Kowal, M., Aldemir, İ. D., Jeftić, A., Memisoglu-Sanlı, A., Najmussaqqib, A., et al. (2022). Big five traits predict stress and loneliness during the covid-19 pandemic: evidence for the role of neuroticism. *Pers. Individ. Differ.* 190, 111531. doi: 10.1016/j.paid.2022.111531
- Irving, R. W. (1985). An efficient algorithm for the "stable roommates" problem. *J. Algorithms* 6, 577–595. doi: 10.1016/0196-6774(85)90033-1
- Javaid, S., Majeed, A., and Afzal, H. (2013). "A reputation management system for efficient selection of disaster management team," in *2013 15th International Conference on Advanced Communications Technology (ICACT)* (PyeongChang: IEEE), 829–834.
- Kamphuis, W., Gaillard, A. W., and Vogelaar, A. L. (2011). The effects of physical threat on team processes during complex task performance. *Small Group Res.* 42, 700–729. doi: 10.1177/1046496411407522
- Kanki, B. G., Folk, V. G., and Irwin, C. M. (1991). Communication variations and aircrew performance. *Int. J. Aviat. Psychol.* 1, 149–162. doi: 10.1207/s15327108ijap0102\_5
- Kanki, B. G., Lozito, S., and Foushee, H. C. (1989). "Communication indices of crew coordination." in *Aviation, Space, and Environmental Medicine*. (Aerospace Medical Assn).
- Kawatsuma, S., Fukushima, M., and Okada, T. (2012). Emergency response by robots to fukushima-daiichi accident: summary and lessons learned. *Ind. Robot.* 39, 428–435. doi: 10.1108/01439911211249715
- Kennedy, B., Curtis, K., and Waters, D. (2014). The personality of emergency nurses: is it unique? *Aust. Emerg. Nurs. J.* 17, 139–145. doi: 10.1016/j.aenj.2014.07.002
- Kichuk, S. L., and Wiesner, W. H. (1997). The big five personality factors and team performance: implications for selecting successful product design teams. *J. Eng. Technol. Manag.* 14, 195–221. doi: 10.1016/S0923-4748(97)00010-6
- Kincaid, J. P., Donovan, J., and Pettitt, B. (2003). Simulation techniques for training emergency response. *Int. J. Emerg. Manag.* 1, 238–246. doi: 10.1504/IJEM.2003.003300
- King, G. (2002). Crisis management team effectiveness: a closer examination. *J. Bus. Ethics* 41, 235–249. doi: 10.1023/A:1021200514323
- Knuth, D. (2021). Steelcraigames. Available online at: <https://twitter.com/SteelCrateGames>
- Kozlowski, S. W., and Klein, K. J. (2000). *A Multilevel Approach to Theory and Research in Organizations: Contextual, Temporal, and Emergent Processes*. Jossey-Bass.
- Kretzschmar, M. E., Ashby, B., Fearon, E., Overton, C. E., Panovska-Griffiths, J., Pellis, L., et al. (2022). Challenges for modelling interventions for future pandemics. *Epidemics* 38, 100546. doi: 10.1016/j.epidem.2022.100546
- Krumm, S., Kanthak, J., Hartmann, K., and Hertel, G. (2016). What does it take to be a virtual team player? the knowledge, skills, abilities, and other characteristics required in virtual teams. *Hum. Perform.* 29, 123–142. doi: 10.1080/08959285.2016.1154061
- Lacoursiere, R. B. (1980). *The Life Cycle of Groups: Group Developmental Stage Theory*. New York, NY: Human Sciences Press.

- Landgren, J., and Nulden, U. (2007). "A study of emergency response work: patterns of mobile phone interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. (San Jose, CA), 1323–1332.
- Lavrakas, P. J. (2008). *Encyclopedia of Survey Research Methods*. Sage Publications.
- Leach, L. S., and Mayo, A. M. (2013). Rapid response teams: qualitative analysis of their effectiveness. *Am. J. Crit. Care* 22, 198–210. doi: 10.4037/ajcc2013990
- Leach, M., MacGregor, H., Ripoll, S., Scoones, I., and Wilkinson, A. (2022). Rethinking disease preparedness: incertitude and the politics of knowledge. *Crit. Public Health* 32, 82–96. doi: 10.1080/09581596.2021.1885628
- Lee, D. H., and Jung, T.-P. (2020). A virtual reality game as a tool to assess physiological correlations of stress. *arXiv preprint arXiv:2009.14421*. doi: 10.48550/arXiv.2009.14421
- Lee, M.-C. (2020). Free the data from the birdcage: Opening up data and crowdsourcing activism in taiwan. *PoLAR* 43, 247–264. doi: 10.1111/plar.12371
- London, C. M. (1998). *Dunant's Dream: War, Switzerland and the History of the Red Cross*. Taylor and Francis.
- Longstaff, P. H., and Yang, S.-U. (2008). Communication management and trust: their role in building resilience to "surprises" such as natural disasters, pandemic flu, and terrorism. *Ecol. Soc.* 13, 130103. doi: 10.5751/ES-02232-130103
- Lykourantzou, I., Antoniou, A., Naudet, Y., and Dow, S. P. (2016). "Personality matters: balancing for personality types leads to better outcomes for crowd teams," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing* (New York, NY: Association for Computing Machinery), 260–273.
- Lykourantzou, I., Vergados, D. J., Papadaki, K., and Naudet, Y. (2013). "Guided crowdsourcing for collective work coordination in corporate environments," in *International Conference on Computational Collective Intelligence* (Springer), 90–99.
- Ma, J., Peng, Y., and Wu, B. (2021). Challenging or hindering? the roles of goal orientation and cognitive appraisal in stressor-performance relationships. *J. Organ. Behav.* 42, 388–406. doi: 10.1002/job.2503
- Majchrzak, A., and More, P. H. (2011). Emergency! web 2.0 to the rescue! *Commun. ACM* 54, 125–132. doi: 10.1145/1924421.1924449
- Mammadov, S. (2022). Big five personality traits and academic performance: a meta-analysis. *J. Pers.* 90, 222–255. doi: 10.1111/jopy.12663
- Marc Cieslak, T. G. (2022). *Ukraine: How Crowdsourcing Is Rescuing People From the War Zone*. Available online at: <https://www.bbc.com/news/technology-60785339>. (accessed May 01, 2022).
- Mark, G., Kun, A. L., Rintel, S., and Sellen, A. (2022). Introduction to this special issue: the future of remote work: responses to the pandemic. *Hum. Comput. Interact.* 1–7. doi: 10.1080/07370024.2022.2038170. [Epub ahead of print].
- Martella, C., Li, J., Conrado, C., and Vermeeren, A. (2017). On current crowd management practices and the need for increased situation awareness, prediction, and intervention. *Saf. Sci.* 91, 381–393. doi: 10.1016/j.ssci.2016.09.006
- Martins, L. L., and Shalley, C. E. (2011). Creativity in virtual work: effects of demographic differences. *Small Group Res.* 42, 536–561. doi: 10.1177/1046496410397382
- Masthoff, J. (2004). "Group modeling: selecting a sequence of television items to suit a group of viewers," in *Personalized Digital Television* (Dordrecht: Springer), 93–141.
- Matsumoto, D., and Juang, L. (2016). *Culture and Psychology*. Boston, MA: Cengage Learning.
- McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *J. Pers. Soc. Psychol.* 52, 1258. doi: 10.1037/0022-3514.52.6.1258
- McKinney Jr, E. H., Barker, J. R., Davis, K. J., and Smith, D. (2005). How swift starting action teams get off the ground: what united flight 232 and airline flight crews can tell us about team communication. *Manag. Commun. Q.* 19, 198–237. doi: 10.1177/0893318905278539
- McManus, I., Keeling, A., and Paice, E. (2004). Stress, burnout and doctors' attitudes to work are determined by personality and learning style: a twelve year longitudinal study of uk medical graduates. *BMC Med.* 2, 29. doi: 10.1186/1741-7015-2-29
- McMichael, M., Beverly, M., Noon, J., Patterson, T., and Webb, G. R. (1999). *Role Improving Under Conditions of Uncertainty: A Classification of Types*. Disaster Research Center.
- Mendonça, D. (2007). Decision support for improvisation in response to extreme events: learning from the response to the 2001 world trade center attack. *Decis Support. Syst.* 43, 952–967. doi: 10.1016/j.dss.2005.05.025
- Mitchell, S. S., and Lim, M. (2018). Too crowded for crowdsourced journalism: Reddit, portability, and citizen participation in the syrian crisis. *Can. J. Commun.* 43, a3377. doi: 10.22230/cjc.2019v44n3a3377
- Mohammed, S., and Angell, L. C. (2003). Personality heterogeneity in teams: Which differences make a difference for team performance? *Small Group Res.* 34, 651–677. doi: 10.1177/1046496403257228
- Muethel, M., Gehrlein, S., and Hoegl, M. (2012). Socio-demographic factors and shared leadership behaviors in dispersed teams: implications for human resource management. *Hum. Resour. Manag.* 51, 525–548. doi: 10.1002/hrm.21488
- Muhren, W. J., van de Walle, B. A., Muhren, W. J., and Van de Walle, B. (2010). Sense-making and information management in emergency response. *Bull. Am. Soc. Inf. Sci. Technol.* 36, 30–33. doi: 10.1002/bult.2010.1720360509
- Myers, I. B. (1962). *The Myers-Briggs Type Indicator: Manual*. Consulting Psychologists Press
- National Research Council. (2015). *Enhancing the Effectiveness of Team Science*. National Academic Press.
- Neuman, G. A., Wagner, S. H., and Christiansen, N. D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group Organ. Manag.* 24, 28–45. doi: 10.1177/1059601199241003
- Normark, M. (2002). "Sense-making of an emergency call: possibilities and constraints of a computerized case file," in *Proceedings of the Second Nordic Conference on Human-Computer Interaction*. (Aarhus), 81–90.
- Odo, C., Masthoff, J., and Beacham, N. (2019a). "Group formation for collaborative learning," in *International Conference on Artificial Intelligence in Education*. (Chicago, IL: Springer), 206–212.
- Odo, C., Masthoff, J., and Beacham, N. A. (2019b). "Adapting online group formation to learners' conscientiousness, agreeableness and ability," in *SLLL@ AIED*. (Chicago, IL), 1–7.
- Okolloh, O. (2009). Ushahidi, or 'testimony': web 2.0 tools for crowdsourcing crisis information. *Participat. Learn. Action* 59, 65–70.
- O'Leary, M. B., and Mortensen, M. (2010). Go (con) figure: Subgroups, imbalance, and isolates in geographically dispersed teams. *Organ. Sci.* 21, 115–131. doi: 10.1287/orsc.1090.0434
- Olson, J. S., and Kellogg, W. A. (2014). *Ways of Knowing in HCI, Vol. 2*. London: Springer.
- O'Neill, T. A., and Allen, N. J. (2011). Personality and the prediction of team performance. *Eur. J. Pers.* 25, 31–42. doi: 10.1002/per.769
- Pajonk, F.-G., Andresen, B., Schneider-Axmann, T., Teichmann, A., Gärtner, U., Lubda, J., et al. (2011). Personality traits of emergency physicians and paramedics. *Emerg. Med. J.* 28, 141–146. doi: 10.1136/emj.2009.083311
- Palen, L., Hiltz, S. R., and Liu, S. B. (2007). Online forums supporting grassroots participation in emergency preparedness and response. *Commun. ACM* 50, 54–58. doi: 10.1145/1226736.1226766
- Perez, A. J., and Zeadally, S. (2019). A communication architecture for crowd management in emergency and disruptive scenarios. *IEEE Commun. Mag.* 57, 54–60. doi: 10.1109/MCOM.2019.1800626
- Petersson, M., Randall, D., and Helgeson, B. (2004). Ambiguities, awareness and economy: a study of emergency service work. *Comput. Support. Cooperative Work* 13, 125–154. doi: 10.1023/B:COSU.0000045707.37815.d1
- Pettet, G., Baxter, H., Vazirizade, S. M., Purohit, H., Ma, M., Mukhopadhyay, A., et al. (2022). Designing decision support systems for emergency response: Challenges and opportunities. *arXiv preprint arXiv:2202.11268*. doi: 10.48550/arXiv.2202.11268
- Pfaff, M. S. (2012). Negative affect reduces team awareness: the effects of mood and stress on computer-mediated team communication. *Hum. Factors* 54, 560–571. doi: 10.1177/0018720811432307
- Pittenger, D. J. (2005). Cautionary comments regarding the myers-briggs type indicator. *Consult. Psychol. J.* 57, 210. doi: 10.1037/1065-9293.57.3.210
- Poblet, M., García-Cuesta, E., and Casanovas, P. (2013). "Crowdsourcing tools for disaster management: a review of platforms and methods," in *International Workshop on AI Approaches to the Complexity of Legal Systems*. (Heidelberg: Springer), 261–274.
- Potts, L. (2013). *Social Media in Disaster Response: How Experience Architects can Build for Participation*. New York, NY: Routledge.

- Quarantelli, E. L. (1988). Disaster crisis management: a summary of research findings. *J. Manag. Stud.* 25, 373–385. doi: 10.1111/j.1467-6486.1988.tb00043.x
- Rammstedt, B., and John, O. P. (2007). Measuring personality in one minute or less: a 10-item short version of the big five inventory in english and german. *J. Res. Pers.* 41, 203–212. doi: 10.1016/j.jrp.2006.02.001
- Reuter, C., Ludwig, T., and Pipek, V. (2014). Ad-hoc participation in situation assessment: supporting mobile collaboration in emergencies. *ACM Trans. Comput. Hum. Interact.* 21, 1–26. doi: 10.1145/2651365
- Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., and Laredo, J. A. (2013). Crisistracker: crowdsourced social media curation for disaster awareness. *IBM J. Res. Dev.* 57, 4–1. doi: 10.1147/JRD.2013.2260692
- Rolland, J.-P. (2002). “The cross-cultural generalizability of the five-factor model of personality,” in *The Five-Factor Model of Personality Across Cultures. International and Cultural Psychology Series*, eds R. R. McCrae and J. Allik (Boston, MA: Springer).
- Ruef, M., Aldrich, H. E., and Carter, N. M. (2003). The structure of founding teams: homophily, strong ties, and isolation among us entrepreneurs. *Am. Sociol. Rev.* 68, 195–222. doi: 10.2307/1519766
- Sabo, R., and Rajčani, J. (2017). Designing the database of speech under stress. *J. Linguist. Jazykovedný Casopis* 68, 326–335. doi: 10.1515/jazcas-2017-0042
- Salas, E., Driskell, J. E., and Hughes, S. (1996). *The Study of Stress and Human Performance. Stress and human performance* (A 97-27090 06-53). Mahwah, NJ: Lawrence Erlbaum Associates.
- Salas, E., Tannenbaum, S. I., Kozlowski, S. W., Miller, C. A., Mathieu, J. E., and Vessey, W. B. (2015). Teams in space exploration: a new frontier for the science of team effectiveness. *Curr. Dir. Psychol. Sci.* 24, 200–207. doi: 10.1177/0963721414566448
- Salehi, N., McCabe, A., Valentine, M., and Bernstein, M. (2017). “Huddler: Convening stable and familiar crowd teams despite unpredictable availability,” in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1700–1713.
- Schilpzand, M., Herold, D., and Shalley, C. (2010). Members’ openness to experience and teams’ creative performance. *Small Group Res.* 41, 56–76. doi: 10.1177/1046496410377509
- Schmidt, A., Wolbers, J., Ferguson, J., and Boersma, K. (2018). Are you ready2help? conceptualizing the management of online and onsite volunteer convergence. *J. Conting. Crisis Manag.* 26, 338–349. doi: 10.1111/1468-5973.12200
- Schneider, R. O. (2011). Climate change: an emergency management perspective. *Disaster Prevent. Manag.* 20, 53–62. doi: 10.1108/09653561111111081
- Sen, A. (1986). Social choice theory. *Handbook Math. Econ.* 3, 1073–1181. doi: 10.1016/S1573-4382(86)03004-7
- Senot, C., Kostadinov, D., Bouzid, M., Picault, J., Aghasaryan, A., and Bernier, C. (2010). “Analysis of strategies for building group profiles,” in *International Conference on User Modeling, Adaptation, and Personalization* (Heidelberg: Springer), 40–51.
- Sharma, R. R., Chawla, S., and Karam, C. M. (2021). “Global gender gap index: world economic forum perspective,” in *Handbook on Diversity and Inclusion Indices* (Edward Elgar Publishing).
- Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. New York, NY: Penguin.
- Smirnov, A., Levashova, T., and Shilov, N. (2011). “Ubiquitous computing in emergency: role-based situation response based on self-organizing resource network,” in *2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (Miami Beach, FL: IEEE), 94–101.
- Smith, K. A., Dennis, M., Masthoff, J., and Tintarev, N. (2019). A methodology for creating and validating psychological stories for conveying and measuring psychological traits. *User Model Useradapt Interact.* 29, 573–618. doi: 10.1007/s11257-019-09219-6
- Song, Z., Zhang, H., and Dolan, C. (2020). Promoting disaster resilience: operation mechanisms and self-organizing processes of crowdsourcing. *Sustainability* 12, 1862. doi: 10.3390/su12051862
- Stachowski, A. A., Kaplan, S. A., and Waller, M. J. (2009). The benefits of flexible team interaction during crises. *J. Appl. Psychol.* 94, 1536. doi: 10.1037/a0016903
- Stehr, N. (2001). *The Fragility of Modern Societies: Knowledge and Risk in the Information Age*. London: Sage.
- Stein, R., and Swan, A. B. (2019). Evaluating the validity of myers-briggs type indicator theory: a teaching tool and window into intuitive psychology. *Soc. Pers. Psychol. Compass* 13, e12434. doi: 10.1111/spc3.12434
- Stephens, K. K., and Robertson, B. W. (2022). 12 social media platforms and broader participation in crisis communication. *Soc. Media Crisis Commun.* 2nd edition. 12. doi: 10.4324/9781003043409-18
- Stuster, J. W. (2011). *Bold Endeavors: Lessons From Polar and Space Exploration*. Annapolis, MD: Naval Institute Press.
- Su, B.-C., Widjaja, A. E., and Chen, J. V. (2012). “Stress in virtual team vs face-to-face team: Is working in virtual team more stressful than face-to-face team?” *Conference paper. Consultation*.
- Subramaniam, C., Ali, H., and Shamsudin, F. M. (2010). Understanding the antecedents of emergency response: a proposed framework. *Disaster Prevent. Manag.* 19, 571–581. doi: 10.1108/09653561011091904
- Tarafdar, M., and Stich, J.-F. (2021). “Virtual work, technology and wellbeing,” in *The SAGE Handbook of Organizational Wellbeing* (London: SAGE), 159–169.
- Thompson, E. R. (2008). Development and validation of an international english big-five mini-markers. *Pers. Individ. Dif.* 45, 542–548. doi: 10.1016/j.paid.2008.06.013
- Thoms, P., Moore, K. S., and Scott, K. S. (1996). The relationship between self-efficacy for participating in self-managed work groups and the big five personality dimensions. *J. Organ. Behav.* 17, 349–362. doi: 10.1002/(SICI)1099-1379(199607)17:4<andlt;349::AID-JOB756>andgt;3.0.CO;2-3
- Thorndike, R. L. (1982). *Applied Psychometrics*. Boston, MA: Houghton Mifflin.
- Tuckman, B. W., and Jensen, M. A. C. (1977). Stages of small-group development revisited. *Group Organ. Stud.* 2, 419–427. doi: 10.1177/105960117700200404
- Van de Ven, A. H., Delbecq, A. L., and Koenig Jr, R. (1976). Determinants of coordination modes within organizations. *Am. Sociol. Rev.* 41, 322–338. doi: 10.2307/2094477
- van de Water, H., Ahaus, K., and Rozier, R. (2008). Team roles, team balance and performance. *J. Manag. Dev.* 27, 499–512. doi: 10.1108/02621710810871817
- Velev, D., and Zlateva, P. (2012). Use of social media in natural disaster management. *Int. Proc. Econ. Dev. Res.* 39, 41–45.
- Vivacqua, A. S., and Borges, M. R. (2012). Taking advantage of collective knowledge in emergency response systems. *J. Netw. Comput. Appl.* 35, 189–198. doi: 10.1016/j.jnca.2011.03.002
- Wagner, S. L., Martin, C. A., and McFee, J. A. (2009). Investigating the “rescue personality”. *Traumatology* 15, 5–12. doi: 10.1177/1534765609338499
- Waller, M. J., Gupta, N., and Giambatista, R. C. (2004). Effects of adaptive behaviors and shared mental models on control crew performance. *Manag. Sci.* 50, 1534–1544. doi: 10.1287/mnsc.1040.0210
- Wauben, L., Dekker-van Doorn, C., Van Wijngaarden, J., Goossens, R., Huijsman, R., Klein, J., et al. (2011). Discrepant perceptions of communication, teamwork and situation awareness among surgical team members. *Int. J. Quality Health Care* 23, 159–166. doi: 10.1093/intqhc/mzq079
- Weick, K. E. (1993). The collapse of sensemaking in organizations: the mann gulch disaster. *Administrat. Sci. Q.* 38, 628–652. doi: 10.2307/2393339
- Weisberg, Y. J., DeYoung, C. G., and Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the big five. *Front. Psychol.* 2, 178. doi: 10.3389/fpsyg.2011.00178
- Wildman, J. L., Shuffler, M. L., Lazzara, E. H., Fiore, S. M., Burke, C. S., Salas, E., et al. (2012). Trust development in swift starting action teams: a multilevel framework. *Group Organ. Manag.* 37, 137–170. doi: 10.1177/1059601111434202
- Worchel, S., and Shackelford, S. L. (1991). Groups under stress: the influence of group structure and environment on process and performance. *Pers. Soc. Psychol. Bull.* 17, 640–647. doi: 10.1177/0146167291176006
- Yang, L., Prasanna, R., and King, M. (2009). On-site information systems design for emergency first responders. *J. Inf. Technol. Theory Appl.* 10, 2.
- Yeo, J., Knox, C. C., and Jung, K. (2018). Unveiling cultures in emergency response communication networks on social media: following the 2016 louisiana floods. *Quality Quantity* 52, 519–535. doi: 10.1007/s11135-017-0595-3
- Yu, T., Sengul, M., and Lester, R. H. (2008). Misery loves company: the spread of negative impacts resulting from an organizational crisis. *Acad. Manag. Rev.* 33, 452–472. doi: 10.5465/amr.2008.31193499
- Yuan, F., and Liu, R. (2018). Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: hurricane matthew case study. *Int. J. Disaster Risk Reduct.* 28, 758–767. doi: 10.1016/j.ijdr.2018.02.003

- Zhang, H. (2022). A real-time optimisation for disaster-relief distribution in heterogeneous crowdsourcing. *Impact* 2022, 23–25. doi: 10.21820/23987073.2022.1.23
- Zhang, Y.-L., and Lu, C.-Q. (2009). Challenge stressor-hindrance stressor and employees work-related attitudes, and behaviors: the moderating effects of general self-efficacy. *Acta Psychol. Sin.* 41, 501. doi: 10.3724/SP.J.1041.2009.00501
- Zijlstra, F. R., Waller, M. J., and Phillips, S. I. (2012). Setting the tone: early interaction patterns in swift-starting teams as a predictor of effectiveness. *Eur. J. Work Organ. Psychol.* 21, 749–777. doi: 10.1080/1359432X.2012.690399

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Vinella, Odo, Lykourentzou and Masthoff. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.