Crowdsourcing User-Centered Teams

Federica Lucia Vinella

Utrecht University

SIKS Dissertation Series No. 2024-21
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

ISBN:	978-90-393-7691-1
Printed by:	ProefschriftMaken
Cover by:	Federica Lucia Vinella
Digital version:	available at `https://dspace.library.uu.nl/`.

This dissertation was printed in limited volume on 100% recycled paper in an effort to minimize the enviromental footprint.

Crowdsourcing User-Centered Teams

Crowdsourcing van gebruikersgerichte teams

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
maandag 17 juni 2024 des ochtends te 10.15 uur

door

Federica Lucia Vinella

geboren op 16 februari 1990 te Putignano, Italië

Promotoren:

Prof. dr. J.F.M. Masthoff, Utrecht University

Copromotoren:

Dr. I. Lykourentzou, Utrecht University

Beoordelingscommissie:

Prof. dr. H.L. Hardman, Utrecht University

Prof, dr. M. Tkalcic, University of Primorska

Prof. dr. J. Vassileva, University of Saskatchewan

Prof. dr. R.C. Veltkamp, Utrecht University

Prof. dr. A. Volk, Utrecht University

Supervisor:

Prof. dr. J.F.M. Masthoff, Utrecht University

Co-Supervisor:

Dr. I. Lykourentzou, Utrecht University

Assessment Committee:

Prof. dr. H.L. Hardman, Utrecht University

Prof, dr. M. Tkalcic, University of Primorska

Prof. dr. J. Vassileva, University of Saskatchewan

Prof. dr. R.C. Veltkamp, Utrecht University

Prof. dr. A. Volk, Utrecht University

Contents

1

Introduction

Online work marketplaces like Fiverr [210], Upwork [462], and Prolific [469] are becoming increasingly popular for companies and individuals to outsource tasks such as software and project development. The number of workers hired depends on the project's size, type, and budget. For larger, more complex projects, crowdsourcing platforms are used. These platforms connect companies with a vast pool of online workers– also known as *crowd workers*. Amazon Mechanical Turk is one of the first such platforms, and Howe et al. [250] is credited with coining the term *crowdsourcing*. This digital labour marketplace initially offered companies the opportunity to hire vast numbers of workers for a limited budget and time and for often repetitive and easy-to-execute tasks known as *microtasks*. Over the years, crowdsourcing has evolved into a dynamic and multifaceted umbrella term where various kinds of digital contributions occur. Nowadays, crowdsourcing is also collaborative, open source, and innovative – thanks to platforms such as Kaggle – for machine learning and data science-based projects [54] and OpenIDEO – for outsourcing projects with a social impact [316]. Whilst assembling good teams is already hard in a traditional setting, this online crowdsourcing setting poses even more challenges. For instance, ad-hoc crowdsourced teams may be very large and geographically dispersed, have no well-defined task, may lack a leader, and members may have no clear roles.

This thesis proposes applying User-Centered Design (UCD) principles to enhance collaborative crowdsourcing systems, improving teamwork, team formation, and overall efficiency. While micro-tasks like captcha and data annotation remain ubiquitous in crowdsourcing [334], there has been a growing demand for more sophisticated and collaborative tasks that necessitate diverse and skilled workforces. For instance, open innovation challenges, collaborative problem-solving tasks, and large-scale software development projects require individuals with different expertise to work together effectively [289]. The shift towards remote work, accelerated by the COVID-19 pandemic [214], has further emphasized the importance of virtual collaboration, leading to the growth of diverse, dynamic, and ill-defined workspaces [270]. As a result, divergent thinking and teamwork have become essential drivers of innovation. Companies like NASA, Procter & Gamble, and Netflix increasingly outsource tasks to diverse crowds to yield novel ideas and comprehensive solutions [256].

Simultaneously, advances in AI have enabled the automation of tasks such as playing chess, real-time translation, and even diagnosing medical conditions [16], liberating workers from repetitive and narrowly defined jobs. This has led to a growing emphasis on corporate social responsibility and purpose-driven employment, with companies such as Patagonia and Salesforce successfully boosting profits through worker empowerment [237]. However, some market-driven crowdsourcing systems often encounter challenges restricting crowd workers' autonomy and agency.

This thesis applies the concept of *User-Centered Design* to collaborative crowdsourcing systems by adapting UCD principles that have traditionally guided the development of user-friendly interfaces, such as consistency, flexibility, and feedback. The main contribution of this research lies in incorporating user needs and characteristics into computational solutions for team formation, teamwork, and team roles within crowdsourcing platforms. This approach aims to empower users, enabling them to influence decisions and foster more inclusive, multi-disciplinary remote collaboration.

By integrating user needs and characteristics into market-driven crowdsourcing systems, this thesis proposes a novel research direction to surpass rigid workflows that offer workers little autonomy. Through a collection of experimental findings, our research intends to lead to more dynamic team formation algorithms considering individual preferences, communication styles, and expertise, ultimately fostering effective and satisfying collaborations and better performance. Finally, this research contributes to collaborative crowd systems' sustainable and scalable design centred on users' needs, behaviours, and characteristics.

1.1. Motivation

This thesis focuses on the individuals participating in crowdsourcing tasks, commonly referred to as *crowd workers*. In particular, it examines collaborative crowdsourcing, where workers join forces to complete tasks such as open innovation challenges, software development, or content creation. However, some crowdsourcing systems overlook the users' characteristics, preferences, viewpoints, needs, and differences, which are crucial in shaping effective teamwork. Furthermore, these systems often neglect hidden attributes like personality types and communication styles, vital for successful collaboration [332]. The primary objective of this thesis is to address these issues and develop a more user-centred approach to crowdsourcing that accounts for crowd workers' diverse characteristics, preferences, and needs, thereby fostering productive collaboration.

In some cases, crowdsourcing systems permit users to self-organize without supervision over the team formation process. As a result, the emerging teams are frequently disconnected from the crowdsourcing platform, with users seeking connections and collaborators through alternative channels such as social media. Such self-organized teams, neither mediated nor monitored by the crowdsourcing system, might exhibit limited diversity of opinions and characteristics [415], potentially giving rise to echo chambers and social homophily that could stifle diverse thinking [208].

Additionally, specific self-organized crowdsourcing collaborative spaces might lack coordination, which becomes particularly critical when projects and users join the system sequentially and at different times. In these scenarios, recurring patterns may include preferential attachment, where the rich get richer, and segregation, which can adversely affect minority groups and hinder radical innovation [138]. However, it is essential to recognize that other factors, such as the nature of the task or the participants' skills, could also contribute to the lack of coordination.

To overcome these challenges, the thesis explores the development of a user-centred approach to collaborative crowdsourcing that considers the diverse needs, preferences, and characteristics of crowd workers. By doing so, it aims to create more effective, inclusive, and satisfying collaborations, ensuring that the system is tailored to the unique requirements of its users. Through a set of studies, we propose tackling some challenges of collaborative crowdsourcing from the angle of users and diversity. By thoroughly examining literature, user behaviours, interfaces, network structures, and algorithms, this work presents design guidelines and methods to develop a user-centric collaboration model that prioritizes teams from the perspective of system users.

The thesis comprises multiple studies investigating factors influencing cooperation, including personality traits, communication styles, cultural backgrounds, and demographics. The aim is to create intelligent crowd systems that accommodate individual differences by offering personalization and recommendations. These studies also explore users' preferences and decision-making processes when forming teams for content creation and problem-solving tasks.

In summary, this thesis contributes to developing a user-centred approach to crowdsourcing collaboration to enhance the effectiveness and outcomes of crowdsourcing initiatives.

1.2. Research Questions

This thesis investigates the critical factors and methodologies for developing user-centred crowdsourcing collaborative systems by focusing on the users and their agency. To explore this, we propose the following overarching Research Question.

RQ: What are the critical factors for developing user-centred collaborative crowdsourcing systems that promote engagement, efficiency, and diversity in online crowd teamwork?

The question focuses on two aspects of designing user-centred collaborative crowdsourcing systems, namely i) *critical factors* that affect online team formation and teamwork and ii) *methodologies* that aid with the formation of more diverse and inclusive crowd teams. We divide the thesis into four Research Questions (RQ1, RQ2, RQ3, and RQ4) with corresponding sub-questions.

The first set of Research Questions (RQ1 and sub-questions RQ1.1, RQ1.2) focuses on understanding crowd workers' preferences regarding the disclosure and visibility of profiling attributes within online crowdsourcing platforms. Profiling attributes refer to the personal and professional information that workers may share to facilitate team formation and project allocation, such as skills, experience, and interests [584]. Given that crowd workers often have limited influence over how crowd platforms are designed and managed [357], we deem it crucial to test whether current design assumptions align with their opinions, preferences, and experiences. Specifically, we aim to explore what information about themselves and potential teammates crowd workers would like to be disclosed. We do this by querying their preferences regarding personal and other teammates' information disclosure. To address these considerations, we investigate the following Research Questions:

RQ1: Which personal and professional profile attributes do crowd users prefer to see and show on crowdsourced team formation systems?

In exploring RQ1, our study focused on understanding the personal and professional attributes crowd workers prefer to be displayed in crowdsourced team formation systems. The findings from our analysis are structured around several sub-questions:

RQ1.1: *About themselves*, which personal and professional profile attributes do crowd workers prefer to display on crowdsourced team formation systems?

(a) Which *types* of attributes (surface-, deep-level) are crowd workers *willing to display* about themselves?

(b) Which *types* of attributes (surface-, deep-level) do crowd workers *find useful*[1] to display about themselves?

(c) Are crowd workers *more willing to display* surface- or deep-level attributes about themselves?

(d) Do crowd workers find it *more useful to display* surface- or deep-level attributes about themselves?

(e) Which *individual* attributes are crowd workers *willing to display* about themselves?

(f) Which *individual* attributes do crowd workers *find useful to display* about themselves?

RQ1.2: *About others*, which personal and professional profile attributes do crowd workers prefer to see on crowdsourced team formation systems?

(a) Which *types* of attributes (surface-, deep-level) are crowd workers *willing to see* about others?

[1] Useful here refers to the perceived utility of the disclosure of information in terms of practical worth or applicability within the team formation context.

(b) Which *types* of attributes (surface-, deep-level) do crowd workers *find useful* to see about others?

(c) Are crowd workers *more willing to see* surface- or deep-level attributes about others?

(d) Do crowd workers find it *more useful to see* surface- or deep-level attributes about others?

(e) Which *individual* attributes are crowd workers *willing to see* about others?

(f) Which *individual* attributes do crowd workers *find useful to see* about others?

The second study addresses the potential for discrimination and prejudice when such profiling attributes are openly displayed to form teams online. Recognizing the natural human tendency to seek out similar others—a phenomenon well-documented in social psychology literature [390], there is a valid concern that online platforms facilitating team formation among crowd workers might unintentionally encourage exclusionary behaviours [202]. To counteract these possible adverse outcomes, our research focuses on the role of interface design and digital interventions—collectively called *digital nudging*. Digital nudging is the strategy of shaping the digital environment to subtly encourage users towards certain decisions or behaviours without limiting their freedom of choice [552]. Specifically, we investigate how a limited selection of nudges can influence the choice of teammates, encouraging crowd workers to form more diverse and inclusive teams. Our study's guiding Research Questions (RQ2 and sub-questions) are as follows:

RQ2: What is the impact of digital nudging techniques on promoting diversity in self-assembled crowd project teams?

(a) *(How) does Priming affect the diversity of the members that crowd users select for their team?*

(b) *(How) does displaying Diversity Information (DI) affect the diversity of the members that crowd users select for their teams?*

(c) *(How) does the combination of Priming and Diversity Information (DI) (Priming + DI) affect the diversity of team members that crowd users select for their teams?*

The third set of Research Questions (RQ3 and sub-questions) investigates whether and how individual and team characteristics influence collaborative tasks when these characteristics are not revealed to crowd workers. The impact of personality traits and communication styles is examined in collaborative crowd work, focusing on how these factors affect team dynamics and task outcomes. To explore this, we designed an experiment where crowd workers collaborated on a high-pressure, time-sensitive task without prior knowledge of their teammates' attributes[2]. Before the task, we collected

[2]The choice of the task was based on 1. novelty (there is a knowledge gap in the literature concerning a deep understanding of ad hoc crowd teams' efficacy in disaster scenarios), 2. utility (understanding which characteristics impact crowd teams the most under stress can be instrumental to future disaster management endeavours), 3. cooperative nature (the imperative cooperative aspect of the task ensured that both crowd

data on participants' surface- and deep-level characteristics to understand how these hidden factors influence collaboration in a stressful scenario. RQ3 is as follows:

RQ3: How do personality, communication patterns, and other user characteristics affect online ad hoc teams under pressure in emergency response situations?

This Research Question guides quantitative and qualitative user-based online research to investigate the effects of crowd characteristics, such as personality traits and communication styles, on online ad-hoc teamwork. Crowd workers were randomly paired into teams, and we evaluated how they completed a time-bounded critical task (defusing a fictive bomb) while working together. For the study, we adopted a strictly cooperative activity to ensure that cooperation is preferred over other strategies, such as competition and social loafing (i.e., putting less effort into the task than other teammates) [85]. We divide RQ3 into three sub-questions:

a) *What personality characteristics render high-stake online teams successful?*

b) *Which skills, abilities, or socio-cultural elements must be considered when forming these teams?*

c) *Are there any particular communication patterns that can serve as early signals of effective teamwork under stress?*

Team formation systems for ad-hoc crowd collaborations often use automated methods, such as algorithms or human mediators, to match collaborators [358]. This process is crucial for assembling effective teams, mainly when participants have limited direct interaction and knowledge about the task and the teammates. In light of this, our study aims to explore how crowd workers might approach the task of team formation themselves, particularly when given limited information about potential team members. This final study focuses on three deep-level attributes—Openness to Experience, Conscientiousness, and Ability—each measured at low, medium, and high levels. To investigate this, we conducted an experimental study where crowd workers assumed the role typically played by a team formation system or algorithm. This approach allowed us to directly assess how crowd workers use deep-level attributes in their decision-making processes for team assembly. The Research Questions (RQ4 and sub-questions) guiding our study are as follows:

RQ4: How does the crowd decide on team formation given profiling attributes?

(a) *Does the even distribution of the team members' attributes differ based on the attribute (i.e., Openness to Experience, Conscientiousness, and Ability)?* This Research has a follow-up sub-question if the answer is true. The sub-question regarding potential disparities in attribute levels, namely high and low, is as follows:

i) *Which attribute level is the most evenly distributed?*

workers would contribute to the outcome). More in Chapter 5.

The sub-research Question concerns differences in even distribution between high and low attribute levels.

(b) *Does cohesion differ based on the attribute?*

(c) *Does the team's balance differ based on the attribute?*

1.3. Methodology

The studies presented in this thesis aim to improve collaborative crowdsourcing systems through a user-centred approach by conducting qualitative and quantitative research across four studies with human subjects (see Figure 1.1). The main goal is to investigate the human factors influencing effective collaboration and team formation in crowd-sourcing settings, including surface and deep-level attributes. Participants gave their informed consent before participating in the experiments. In the following sections, we provide an overview of the approaches, metrics, and methods used in the studies to gain insights into enhancing collaboration and team formation in crowdsourcing systems.

Study 1: Understanding Crowd Team Member Profiling Preferences. This study gathers the opinions of crowd participants about exposing and observing information about themselves and other crowd workers on online team formation tools. The analysis was primarily quantitative, as it asked participants to rate (using a five-point Likert scale) a set of profiling attributes according to their willingness to see/show and usefulness in team formation. The study revealed that crowd workers prefer to display and view surface-level attributes, particularly demographics and social-media-related features (e.g., availability, profile photo, etc.). Deep-level attributes related to mental states, beliefs, and political affiliations are less preferred overall. However, personality, opinions, and values are exceptions, being considered valuable in collaborative settings. The results indicate a general preference for disclosing surface-level attributes on online team formation tools. They also demonstrate that certain deep-level traits are considered relevant by crowd workers who are – at least in principle – willing to disclose covert information with the understanding that it may improve team formation. The work is presented in Chapter 3.

Study 2: Nudging Diversity in Crowd Team Formation. In the previous study, we found that crowd workers want to see surface attributes and think these are useful. There is, however, a danger that this may lead to teams that lack in diversity, given people's unconscious biases [415]. This study aimed to see if digital interventions could help crowd workers choose a more diverse set of teammates. We tested this with an online creative project. Since people (and crowd workers alike) often favour teammates similar to themselves [144], and having a variety of people on a team can lead to better and more innovative solutions [632], we investigated if and how different types of nudging interventions could make a difference in encouraging more diversity in teammate selection. For the profiling attributes, we used a combination of surface and deep-level traits found relevant according to the crowd workers (in Chapter 3).

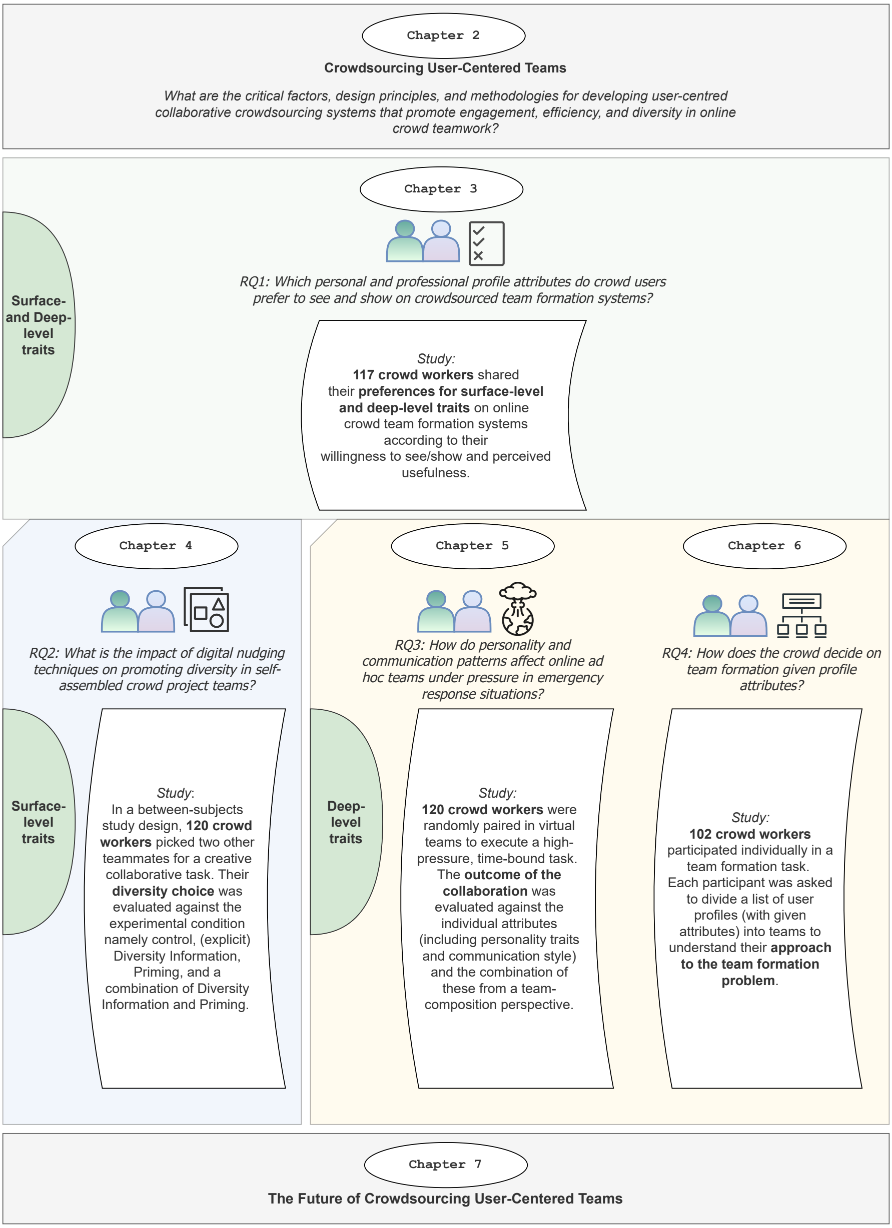


Figure 1.1: Thesis structure. The thesis comprises four studies based on the notion of *User Centred Design* applied to collaborative crowdsourcing, system development, and diversity in the workspace.

We combined this selection of attributes with others commonly used in online team formation systems (e.g., [202]). By setting up a 2x2 factorial analysis, we compared the behaviour of the crowd participants when exposed to four different digital nudging conditions. The study collected demographic information from crowd participants and adjusted the nudging interventions based on this information (e.g., by changing the content of the priming intervention) to recommend teammates within the diversity angle. The work is presented in Chapter 4.

Study 3: Exploring Crowd Team Attributes and Dynamics. Given that specific deep-level profiling attributes such as personality traits were considered relevant by crowd workers in Chapter 3, in this chapter, we assess to what extent a selected set of covert traits relates to (and potentially may impact) teamwork performance during a crowd-sourcing collaborative task. More precisely, this study identified how some deep-level traits, such as personality traits, communication styles and cultural background, impact tasks that require close collaboration and are time-sensitive. In the study, participants defused a bomb in a virtual maze as part of an ad-hoc and randomly assigned team task. The team (comprising two individuals) chatted to discuss the best course of action within a limited time. We gathered information about the crowd workers' personality traits through the BFI-10 inventory (a short version of the Big-5 personality inventory [476]).

At the end of the task, participants filled out a questionnaire regarding the collaboration that helped gather insights into their perception of several aspects of teamwork (cohesion, communication, balance, and satisfaction). The work enabled us to recognize deep-level (and some surface-level) attributes that independently and jointly contribute to crowd teamwork under pressure. In particular, it revealed that teams with higher minimum levels of openness performed better, as did those with a higher frequency of action and response statements (i.e., a coded focused communication style). Furthermore, highly agreeable individuals tended to view their team's performance more positively, even in defeat. Lastly, the study revealed that the effectiveness of and satisfaction with communication patterns were role-dependent, indicating the importance of aligning communication strategies with team roles. We report this work in Chapter 5.

Study 4: Eliciting Crowd's Top-Down Strategies in Team Formation. In the studies above, we understood that surface-level traits are commonly used and accepted on online team formation tasks (and must be accounted for when designing for inclusion and diversity). From the results, we also found that deep-level traits are considered relevant and play a significant role in online teamwork under pressure. Building on these insights, we investigated how crowd participants would use deep-level information to form teams – if they were given the top-down role of the team formation algorithm. For this study, we followed the User-as-Wizard method [380] and asked crowd participants to drag and drop cards representing dummy profiles into four teams. The study aims to understand whether and how crowd workers handle deep-level attributes in team formation. We used various levels of Conscientiousness, Openness to experience, and Ability for the study design. This study showed that crowd workers prefer to distribute deep-level attributes evenly across teams. In particular, a significant finding from this research is the crowd's general tendency to distribute the trait of Openness to Experience more evenly compared to Conscientiousness and Ability. This suggests that the crowd may focus more on specific deep-level characteristics from the same family (i.e., personality traits) when forming teams. The study is presented in Chapter 6.

1.4. Thesis outline

The need for effective crowdsourcing practices has grown significantly in recent years, and researchers have been exploring ways to improve collaborative crowdsourcing systems. The studies presented in this thesis explore novel crowdsourcing approaches through the user-centred design lens. Chapter 2 covers related work, while Chapters 3, 4, 5, and 6 present various studies on user-centered crowd teams.

Chapter 2: Crowdsourcing, Crowd Teams, and User-Centered Design. This chapter introduces the notion of crowdsourcing, teams, and crowd teams. It highlights issues present in crowdsourcing practices and links five user-centred design approaches to developing crowdsourcing systems for collaboration.

Chapter 3: Understanding Crowd Preferences for Team Member Profiling. This chapter presents a study on crowd participants' preferences and opinions regarding profiling characteristics in online crowd team formation. It collects responses from 117 participants about their preferences for profiling attributes related to surface and deep-level traits, demonstrating the importance of a user-centred approach in designing collaborative crowdsourcing systems.

Chapter 4: Leveraging Digital Nudging to Enhance Crowd Team Diversity. This chapter introduces the concept of digital nudging in crowd team formation. A study with 120 crowd participants compares the effects of four digital interventions on a team formation system where participants could choose teammates for an outsourced task. The interventions followed a 2x2 factorial design with two treatments (diversity information and priming conditions) and two settings (none and applied). The results indicate that specific interventions positively impact diversity, while others have neutral or detrimental effects on nudging the crowd towards diverse teammates.

Chapter 5: Exploring Attributes and Dynamics in High-Pressure Crowd Teams. This chapter investigates the characteristics affecting crowdsourced teams in high-pressure, time-sensitive tasks. It presents a study with 120 crowd participants randomly paired to perform a high-pressure task within a limited time, examining the effects of personality traits, communication, and demographic characteristics on team composition and teamwork outcomes.

Chapter 6: The Wisdom of the Crowd in Team Formation. This chapter investigates team formation from the perspective of the crowd. In a user-as-wizard study, 120 crowd participants formed groups by dragging and dropping dummy profiles with information about different levels of personality and ability traits into boxes representing four teams. The study shows that the wisdom of the crowd approach to team formation attempts to distribute resources fairly across teams.

Chapter 7: Implications, Limitations, and Future Directions. This chapter discusses the limitations of the studies and explores recent comparisons from the literature. We also suggest future directions for follow-up studies or system guidelines.

1.5. Contributions to Knowledge

The chapters in this thesis showcase how a user-centred design approach can significantly advance crowdsourced team formation systems. The four user-based studies presented provide a comprehensive analysis of collaborative crowdsourcing systems. By summarizing and formalizing the results of these studies, this thesis offers a set of guidelines to enhance AI-based teamwork in emergency response, recommendations for profiling crowd workers in team formation settings, parameters for automating team formation online, and suggestions for designing more inclusive systems that consciously use user interfaces to encourage diversity choices.

Guidelines for profiling crowd teams. From the results collected in our survey (Chapter 3), we have extrapolated general preferences for specific traits when joining team formation systems for crowd work. Surface-level attributes such as gender, age, and appearance (in the form of a photo) are favoured by the crowd, followed by deep-level traits such as topical interests, availability, and rating. We also emphasize the importance of avoiding the disclosure of sensitive attributes such as political affiliation, ethnicity, and mental states like depression. The results inform system designers when developing tools for crowd team formation.

Recommendations for designing inclusive crowd team formation systems. From the results of a comparative user study (Chapter 4) on nudging the crowd towards more diverse teammate selection, we provide recommendations that we found effective in motivating users towards diversity. Firstly, we suggest applying and experimenting with explicit representations of diversity in UI interventions. In our case, progress bars indicating the overall diversity between users positively drove more diverse choices. We also recommend avoiding subtle or subliminal approaches, such as priming, as they may be sources of misinterpretation and ambiguity that could diminish diversity.

Guidelines for AI support for teams in emergency response. Our study in Chapter 5 provides an evidence-based list of parameters to enhance AI support for crowd teams in emergency response. The first guideline recommends considering specific personality traits (minimum Openness to Experience) during team formation for teamwork under pressure. The second suggests using conversational AI to detect hidden information about the team and leveraging this knowledge by adapting the communication style to the crowd's various roles, personalities, and cultural differences.

Parameters for automated team formation. Through the wisdom of the crowd and observational methods (Chapter 6), we provide a user-centred set of parameters for automating team formation online. In contexts like online education, we identify Ability, Conscientiousness, and Openness to Experience as the three most preferred attributes according to the crowd when assembling teams of learners. Based on the User as Wizard evaluation results, we also suggest that balancing characteristics within and between teams is a preferred strategy that automated algorithms can emulate.

2

Crowdsourcing, Crowd Teams, and User-Centered Design

In this chapter, we introduce the concept of crowdsourcing (Section 2.1) and explore its components (Section 2.2). Then, we discuss research methods used in crowdsourcing studies (Section 2.3) and examine the characteristics affecting teamwork (Section 2.4). We then look at some of the most common types of crowdsourced teams (Section 2.5). Next, we address the limitations and challenges crowdsourcing systems face (Section 2.6). Ultimately, we propose applying five User-Centered Design principles to collaborative crowdsourcing systems research (Section 2.7).

2.1. Crowdsourcing: Weaving Collective Ingenuity

Mark Twain's classic tale of Tom Sawyer provides an early insight into the concept now known as crowdsourcing. Faced with the tedious chore of painting a fence, Tom cleverly enlists the collective efforts of his friends by making the task appear enjoyable and prestigious. This cunning strategy benefits everyone: Aunt Polly receives a perfectly painted fence, the boys enjoy the engaging task, and Tom, a forerunner to today's crowdsourcing platform managers, significantly increases his wealth.

Crowdsourcing has a deeply rooted history spanning centuries. Numerous examples of collaborative problem-solving and innovation exist, such as the 18th-century "Longitude Prize" offered by the British government and the 1919 Planters Peanuts logo design contest. More recently, the 1990s open-source software movement paved the way for internet-based crowdsourcing, exemplified by projects such as the Linux operating system. Although its historical roots run deep, crowdsourcing as an approach is relatively new. Jeff Howe and Mark Robinson introduced it in a Wired Magazine article, capturing the essence of utilizing crowd power for various purposes, from problem-solving to innovation. Crowdsourcing gained traction with the advent of platforms such as Amazon Mechanical Turk in 2005, which connected businesses and individuals to a global workforce for microtask completion. Its popularity surged, and applications broadened to encompass complex problem-solving domains, such as scientific research, urban planning, journalism, and product design. The field's evolution has been characterized by specialized platforms and tools catering to different industries and sectors' unique needs. These platforms allow organizations to access the collective intelligence of diverse, geographically dispersed individuals, promoting creativity, innovation, and collaboration.

Examples of successful crowdsourcing projects include Foldit, a protein-folding game that enhanced scientific understanding of protein structures [290]; Galaxy Zoo, which enables citizens to classify galaxies and contribute to astronomical research [471]; and Kickstarter, a crowdfunding platforCrowdsourcingnds of creative projects [399]. Crowdsourcing faces challenges and limitations, such as quality control, participant motivation, and ethical concerns. Addressing these issues is essential for a balanced perspective on the topic [357].

Artificial intelligence and machine learning have further revolutionized crowdsourcing, with hybrid approaches emerging, which combine human and machine intelligence. These strategies offer new opportunities to capitalize on the strengths of both humans and machines for solving complex problems. As crowdsourcing continues to develop, its potential to drive innovation and address challenges across various domains is anticipated to grow. Technological advancements, organizational structure shifts, and new strategies for harnessing crowd power will likely influence crowdsourcing practices in the future. Furthermore, there is an increased need for creating user-centred crowd teamwork systems based on collaboration and remote communication. This thesis examines this aspect, proposing a series of studies that apply user-centred design principles in crowdsourced collaborative digital environments.

2.2. Crowdsourcing components

Crowdsourcing practices can be classified based on several attributes, including the type of target problem, the nature of collaboration, the crowd, recruitment and compensation, and workflow management [145, 240, 627, 145, 206]. This section further explores these common forms of classification, providing insight into various tasks and collaboration methods (see Table 2.1 for an overall view).

2.2.1. Contribution modalities: Explicit and Implicit Crowdsourcing

Crowdsourcing encompasses two modalities based on the contributors' engagement: *explicit* and *implicit* crowdsourcing. The following paragraphs explore these distinct forms of crowd contribution.

In explicit crowdsourcing, user collaboration is sought for evaluating, sharing, and accomplishing designated tasks [461]. The term "explicit" refers to the process of soliciting information, services, or ideas from a large, diverse group of individuals (the "crowd") through a *direct* appeal for their engagement [145]. With well-defined tasks or projects and explicit instructions outlining expectations, examples include online

Table 2.1: Main crowdsourcing components (nature of contribution, target problem, the crowd, recruitment and compensation, and workflow management) and their sub-components and secondary factors.

Component	**Sub-component**	**Secondary factors**
Nature of contribution	Implicit contribution	The crowd:
		Piggyback (i.e., Exploiting traces that users leave)
		Component providers
	Explicit contribution	The crowd:
		Assistants
		Perspective providers
		Experts
Target problem	Micro task	Cognitive effort
		Quality of impact
		Ease of Execution
	Macro task	
The Crowd	Motivation	Intrinsic:
		Altruism
		Personal achievement
		Extrinsic:
		Self-marketing
		Social status
		Instrumental motivation
		Token compensation
		Market compensation
Recruitment/ compensation	Paid	Flat rate (e.g., request, pay)
		Bonus prize
	Unpaid	Asking for volunteers
		Making pay for services
		Piggyback
Workflow management	Accessibility of peer contribution	Modify
		Assess
		View
		None
	Aggregation of contribution	Integrative
		Selective

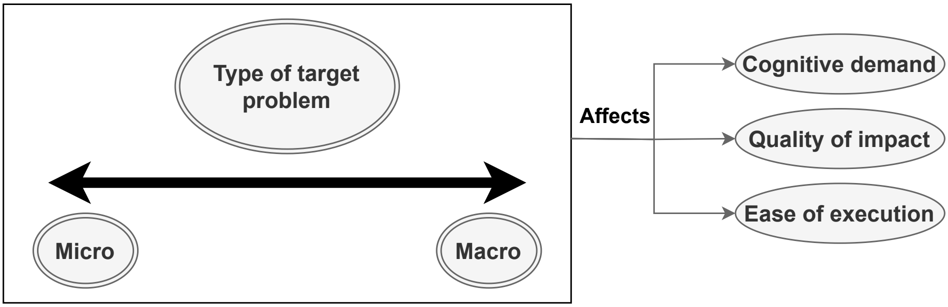


Figure 2.1: The target problem (whether micro or macro) affects key components of crowdsourcing systems, such as the crowd's cognitive effort, the quality of impact, and ease of execution.

surveys, opinion polls, and contests where participants contribute ideas or designs for specific products or services. Companies and organizations may leverage this approach to gather customer feedback, generate novel product concepts, or address targeted problems (e.g.,[320]), offering a cost-effective and efficient means to rapidly accumulate information and ideas from a large pool of individuals.

Implicit crowdsourcing gathers user data without their direct awareness or involvement in a particular project [117]. It involves collecting data or insights from individuals without explicit awareness or active participation in a specific task. This approach collects data by monitoring user behaviour or extracting information from social media platforms. Examples include analyzing search engine queries to discern trends, employing artificial intelligence to scrutinize social media posts, or tracking website traffic patterns to infer user preferences [614]. Market research commonly uses implicit crowdsourcing to gain insights into consumer behaviour and preferences [206]. However, with the increasing popularity of generative AI trained on extensive datasets procured from various online sources [376], this form of crowdsourcing oftentimes raises ethical concerns about privacy and consent. As individuals may remain unaware that their data is being collected and analyzed, careful examination of the moral implications inherent in implicit crowdsourcing practices is necessary.

2.2.2. Target Problem: Micro and Macro Tasks

In crowdsourcing platforms, the "target problem" refers to the specific task or challenge the platform aims to address, usually based on the complexity and scope of the job. Classifying crowdsourcing platforms based on their target problem typically results in two main divisions of labor: *micro-tasking* and *macro-tasking*. In this section, we describe the characteristics of micro and macro tasking and how they differ in cognitive effort, quality of impact, and ease of execution (see Figure 2.1, Table 2.2).

The vast majority of crowdsourcing comprises short tasks executed by individuals without the help of a team. This type of contribution is known as a micro-task. Examples include transcribing speeches into written sentences, spell-checking short paragraphs, sentiment analysis of social media posts, product reviews, transcription of scanned documents, and tumour recognition from lung scans [140, 99, 57]. The micro approach

Table 2.2: Factors and Examples concerning the type of target problem in crowdsourcing. Typically, micro-tasks are less demanding on cognitive effort, quality of impact, and ease of execution than macro tasks.

Factor	Micro-tasking		Macro-tasking	
	Demand	Examples	Demand	Examples
Cognitive effort	Low	Evaluations [18, 466]; Organizational tasks [22]	High	Policymaking [378]; Product development [509]
Quality of Impact	Focused, incremental	Distributed Human Intelligence (HITs) [135]	Significant, broad	Knowledge discovery [450, 526, 306]; Broadcast search [317]
Ease of Execution	Simple, quick	Machine contribution and automation [180]	Complex, collaborative, interdependent	Self-organization and governance [356]

to crowd work pivots around information segmentation and decomposition [356]. It allows users (i.e., crowd workers) to perform small tasks without expert knowledge and for low amounts of money. The crowdsourcing process for micro-tasks involves requesters creating Human Intelligence Tasks (HIT) using their data and a task choice (these are usually short projects that last minutes up to a few hours). The requesters then post these HITs on the crowdsourcing platform, specifying requirements and payment for completion. Workers applying to the publicized HITs perform the assignments and return their submissions to the requester, who receives the desired outcome and compensates crowd workers accordingly.

Contrary to micro-tasking, macro-tasking is a distinctive type of crowdsourcing that addresses complex, wicked, or large-scale challenges that are not readily decomposable or do not have a single, definitive solution [356]. This approach necessitates teamwork, cooperation, and coordination among crowd participants. Examples include designing policies for charitable giving [378] and product development contests [509]. A significant subset of macro-tasking is *crowdsolving*, an integrative and collaborative problem-solving approach in which numerous remote contributors form communities to share ideas, resources, and expertise [10].

Macro-tasking typically demands specialized skill sets, domain proficiency, and a higher degree of autonomy and interdependence among contributors [356]. This work presents unique challenges and opportunities for developing, implementing, and evaluating collaborative systems and processes. To support macro-tasking effectively, researchers are considering the complexities and dynamics of large-scale collaboration, such as flexible coordination mechanisms [358], innovative communication tools [606], and adaptive information-sharing. Moreover, fostering a sense of shared pursuit, trust, and mutual accountability among crowd participants is another factor influencing the outcomes in macro-tasking scenarios.

Cognitive effort

The relationship between macro and micro-tasking in crowdsourcing and the varying levels of *cognitive effort* can be better understood by considering how users are allocated to tasks based on their abilities [113], limits [189], and the requester's requirements [145]. Typically, micro-tasking involves simple, short tasks requiring lower levels of cognitive demand [113] [1]. Examples of low-level cognitive demanding tasks are *evaluations* entailing workers' contribution to quality assessment and other subjective feedback provisions (e.g., evaluating the attractiveness of faces [18] or the adequacy of prices [466]). *Organizational tasks* are another example of moderately demanding micro-tasks as they involve data categorization, information extraction, and content moderation [22]. These tasks may require cognitive effort but are still manageable for individuals without domain-specific expertise. Participants in these tasks might be responsible for labelling images, sorting data, or filtering out irrelevant or inappropriate content. In contrast, macro-tasking encompasses complex, large-scale challenges that often require higher cognitive effort [113]. These tasks may necessitate specialized skill sets [113], proficiency in a given domain [189], or collaboration between crowd participants [493, 555]. By allocating macro-tasks to users with appropriate expertise, the system can benefit from their knowledge and skills, leading to more effective problem-solving and innovation [64, 588]. Examples of high-level cognitive demanding tasks in macro-tasking include designing policies for charitable giving [378], contests for product development [509], and complex problem-solving tasks that require a collaborative and comprehensive approach [356, 354, 358].

Quality of impact

As seen in the previous section, the impact of a contribution varies across tasks [145]. Some are limited in scope, such as editing sentences on Wikipedia, while others have greater outreach and influence, such as drafting policies and rules. Typical examples of low-impact micro-tasks are *Distributed Human Intelligence (HITs)* tasks. They involve short-term, specific tasks that require human intelligence but do not necessarily require domain expertise [135]. HITs include image recognition, data annotation, or categorization [155]. The level of impact for these tasks can vary, but they generally support larger projects or research efforts by processing data or providing feedback. Higher-impact tasks usually belong to macro tasks. Examples are *knowledge discovery* tasks involving solving complex, ambiguous, or wicked problems that require collaboration, specialized skills, and deep domain knowledge [84]. Knowledge discovery crowdsourcing includes health hackathons, data marathons, and crowdsourced research in genetic studies [450], housing [526], or misinformation detection [306]. These tasks have a high impact as they contribute to new knowledge and advancements in various fields and address socially relevant issues [356]. However, knowledge-discovery tasks, which have a high impact, can also take the form of micro-tasks. A notable example is GalaxyZoo, where users use micro-tasking, such as image recognition, to contribute to discovering new scientific knowledge, such as identifying galaxies. Therefore, in this thesis, we suggest separating the low/high impact aspect from the micro/macro axis while considering both factors independently.

[1] However, it's important to note that cognitive load doesn't necessarily correlate directly with expertise requirements. A task can be cognitively challenging without necessitating specialized expertise.

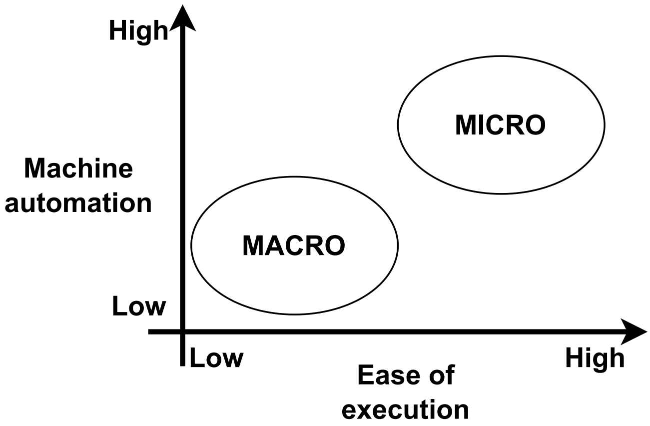


Figure 2.2: The ease of execution of a crowdsourcing task determines the level of machine automation. This usually results in micro tasks having higher ease of execution and automation than macro tasks.

Furthermore, crowdsourcing tasks can have either low or high impact, depending on the nature of the problem and the level of expertise required to solve it. One such example is *broadcast search* tasks, which involve broadcasting requests for information, ideas, or solutions to a large and diverse group of people [8]. In some cases, broadcast search tasks can align with micro-tasking, as they may involve simple tasks or information gathering that a broad audience can complete. On the other hand, there are instances where broadcast search may require more specialized knowledge, diverse perspectives, or collaboration, making it more akin to macro-tasking. Examples of high-impact broadcast search include online forums where users can ask questions and receive answers from a community of experts or enthusiasts [317], as well as social media campaigns where a company or organization requests help finding a solution [436].

Ease of execution

A well-designed crowdsourcing system considers human factors such as mental workload, situation awareness, and skill degradation when allocating tasks to either the crowd or a machine for automation [442]. Assessing the complexity of a job is crucial in determining whether humans or machines (e.g., AI) are better suited to perform it [145]. For example, some jobs are highly challenging for humans (e.g., high-order computations) but straightforward for machines. In contrast, tasks that are intuitive and easy for humans to accomplish (e.g., common sense, empathy, ethical and moral judgment, causality [246]) can be challenging for machines. Furthermore, the complexity of a task often links with the level of autonomy and coordination efforts, also known as community-driven workflows (see Figure 2.2). Macro-tasking crowdsourcing platforms such as Innocentive, Upwork, and OpenIDEO demonstrate this. They manage their contributors with less direct control and surveillance than micro-tasking platforms such as Amazon Mechanical Turk or Prolific [180]. Generally, micro-tasking crowdsourcing tends to rely more on automated and streamlined workflows that depend on computa-

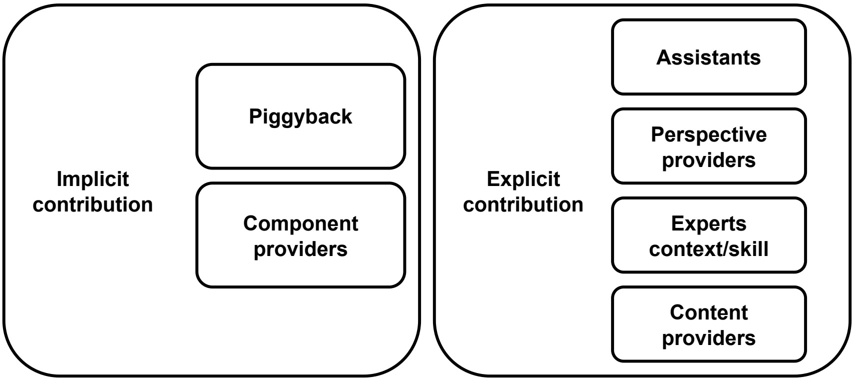


Figure 2.3: Crowd participants' participation is categorized by various naming conventions typically clustered underneath Implicit and explicit contributions.

tional solutions to gather resources. In contrast, macro-tasking crowdsourcing allows human contributors leeway for self-coordination and management [356].

User interfaces (UIs) play a significant role in defining constraints on what crowds and systems can achieve [472]. This aspect is closely related to *machine contribution*, as UI design should consider the distinct modus operandi of humans (e.g., using natural language and fuzzy output to communicate) and machines (e.g., compiling error-free instructions). UIs should facilitate problem-solving by adapting to human needs and ergonomics [145]. Additional criteria for designing UIs that effectively distribute work between humans and machines include assessing automation reliability [59], evaluating the risks associated with decision/action outcomes, and considering the ease of systems integration [442, 168]. Differentiating between difficulty levels helps streamline and optimize the process of assigning crowdsourcing tasks, ensuring that the most appropriate party handles each task. Beyond the type of target problem (micro and macro) and the nature of collaboration (implicit and explicit), crowdsourcing practices depend on several other critical components: participants (or crowd), motivational factors, compensation, and management. In the following sections, we will explore how each element plays a fundamental role in crowdsourcing and how they may differ.

2.2.3. Understanding the Roles and Terminologies of Crowd Participants

The success of crowdsourcing systems relies heavily on the participants, often called the "crowd." This crowd comprises diverse backgrounds, skills, and experts contributing to various tasks and projects (see Figure 2.3). People may participate for several reasons, such as monetary compensation [113], intrinsic motivation, or personal interest in the project [493]. To ensure a crowdsourcing system's quality and effectiveness, attracting and retaining the right crowd is crucial to address the given task [188].

Depending on the nature of their contribution, crowd workers are referred to by different names. In *implicit crowdsourcing*, where participants are unaware that their input is

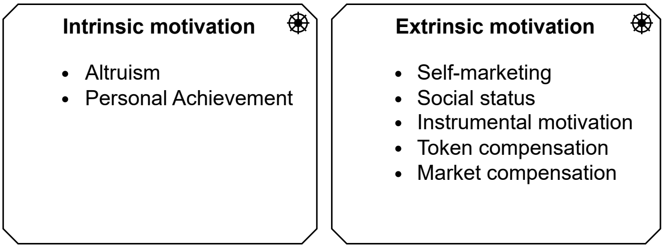


Figure 2.4: Crowdsourcing contribution is triggered by the crowd's intrinsic and extrinsic motivations.

being used for projects, their role is known as *piggybacking* [113]. Piggybacking involves leveraging user traces to generate results, with cookies and search engine queries being common recruitment strategies that capitalize on users' behaviour. Another term for contributors of implicit tasks is *component providers* [145], which refers to how the crowd serves as a component in the target artefact, such as nodes and connections in social networks or community builders. Component providers may not always be aware of their contribution to data collection for crowdsourcing purposes, as their online activities (e.g., posting on social media or reading articles) can be scraped by third-party tools.

The terminology for describing crowd participants in *explicit crowdsourcing* is quite diverse. For micro-tasks, crowd contributors are often referred to as *assistants* [145]. The assistant's role is to help conserve the system (or task) owner's resources, such as time and effort. Managed through a divide-and-conquer approach, the problem is broken down into smaller sub-problems, with the crowd working iteratively until the main task is resolved. Another term for crowd workers in micro-tasks is *perspective providers* [145]. This type of human contribution involves aggregating perspectives, as the combined solutions of multiple contributors are generally more optimal than a single worker's output. Micro crowdsourcing typically aligns with *any individual and most people's tasks* [113], attracting a broad audience without specific qualifications or strict selection criteria [188]. For macro tasks, different names describe the crowd. *Content providers* [145] are a crowd-focused on content creation, such as participatory media, including YouTube videos [28] and Flickr images [275]. Content providers are often motivated intrinsically [493] and tend to be more self-reliant than crowd workers engaged with outsourcing platforms such as Amazon Mechanical Turk for Human Intelligence Tasks (HITs). Macro tasks, which are often ill-defined and wicked [356], require *experts* [113] or *qualified workers* based on criteria such as context and skill sets [188]. These tasks necessitate expertise, unique abilities, or specializations, such as protein folding [204], and geometric packing [263]. By definition, expert tasks make up a smaller portion of crowdsourcing problems and tend to be more expensive [624] than those tasks suitable for any individual or most people [113].

2.2.4. Motivational factors: Exploring the underlying motivations in crowdsourcing

Understanding the motivations behind crowdsourcing work is crucial for attracting and retaining participants, ultimately contributing to a project's success [493, 207]. This section will discuss the diverse motivational factors influencing crowd workers' engagement in various tasks and challenges. Before delving into specific motivational factors, it's important to distinguish between intrinsic and extrinsic motivation.

Intrinsic motivation refers to the internal drive to engage in an activity for its inherent satisfaction and enjoyment.

Extrinsic motivation, on the other hand, stems from external factors, such as rewards or recognition. Both types of motivation can influence an individual's decision to participate in crowdsourcing activities [493]. An overview of the different types of motivational factors is provided in Figure 2.4.

Self-marketing is a powerful motivator for individuals who wish to demonstrate their skills and competencies [330]. For instance, open-source software developers can enhance their professional reputation and potentially secure better employment opportunities by showcasing their expertise through contributions to favored projects. This demonstrates how self-marketing can help individuals build their careers through crowdsourcing platforms.

Social status also plays a significant role in motivating participation [330, 493]. Online communities such as Stack Overflow offer platforms for experts to share their knowledge and gain recognition among their peers. By accumulating points and badges, users can demonstrate their expertise and receive validation from the community, illustrating how social status acts as a driving force for participation.

Instrumental motivation involves pursuing benefits for oneself or a company [330]. In data science competitions hosted by companies such as Kaggle, participants are incentivized with cash prizes and potential job offers. This competitive environment showcases how instrumental motivation encourages individuals to develop solutions to complex problems and, in turn, gain access to valuable resources and networks.

Token compensation, where participants receive small monetary rewards or goods in exchange for their involvement, is another factor that can drive engagement in crowdsourcing [286]. Examples include receiving virtual goods, gift cards, or points that can be redeemed for products or services.

Market compensation, an extension of token compensation, encompasses more substantial financial incentives such as a living wage or salary [330]. In crowdsourcing, this type of motivation can attract skilled professionals or freelancers who rely on their work for financial stability and long-term employment.

Altruism, or the genuine desire to contribute to the welfare of others without expecting personal rewards, can also be a strong motivator in crowdsourcing [528]. For example, the collaborative efforts to map disaster-affected areas through platforms such as OpenStreetMap illustrate how altruistic motivations can drive individuals to help solve societal problems and support nonprofit organizations.

Table 2.3: Recruiting and compensating crowdsourcing contribution is either paid or unpaid or explicit or implicit.

	Paid (flat rate, bonus prize)	Unpaid (volunteering, none)
Explicit contribution	Required (e.g., outsourcing company); Pay (e.g., Amazon Mechanical Turk)	Asking for volunteers (e.g., No-profit)
Implicit contribution		Making pay for services (e.g., Captcha); Piggyback (e.g., social media content)

Lastly, *personal achievement* (including learning motivations) focuses on intrinsic gains such as acquiring new skills, enhancing mastery, boosting competence, and achieving personal fulfilment [330, 493]. In crowdsourcing, these motivations often drive participants to join idea contests or tackle challenging tasks, as they offer valuable learning experiences and feedback opportunities.

2.2.5. Strategies for Recruiting and Compensating Crowd Participants

The recruitment and compensation of crowd participants in crowdsourcing initiatives can significantly influence the success and sustainability of these projects. Various strategies for recruiting and compensating crowd participants can be broadly categorized into paid and unpaid approaches, each with unique advantages and challenges [145] (see Table 2.3).

Explicit recruitment strategies are often paid, requiring consent from participants to engage in specific tasks. One such approach is the *requiring users* strategy, which involves employing internally managed or in-house employees for ad hoc team projects, sometimes in collaboration with outsourced contributors from crowdsourcing platforms. Companies commonly use this approach to develop tools and answer surveys, combining their in-house expertise with external skills [588]. Another paid strategy, *paying users*, involves outsourcing tasks to various contributors on platforms such as Amazon Mechanical Turk. These paid users receive compensation for their work, helping to attract and retain participants. The type and payment amount can significantly impact the quality and quantity of work submitted.

Payment methods are crucial in determining how contributors receive compensation for their work. Corney et al. [113] differentiate between *rewarded contribution at a flat rate* and *with a bonus or prize (i.e., success-based)*. The former provides direct financial incentives, such as payments for completed tasks. It is commonly used in micro-tasking projects where participants perform simple, short tasks for low amounts of money. The latter involves competitions and awards, where participants compete for a prize or recognition, thus encouraging innovation and creative thinking among the crowd. Combinations also occur where crowd participants receive a flat rate plus a bonus on performance.

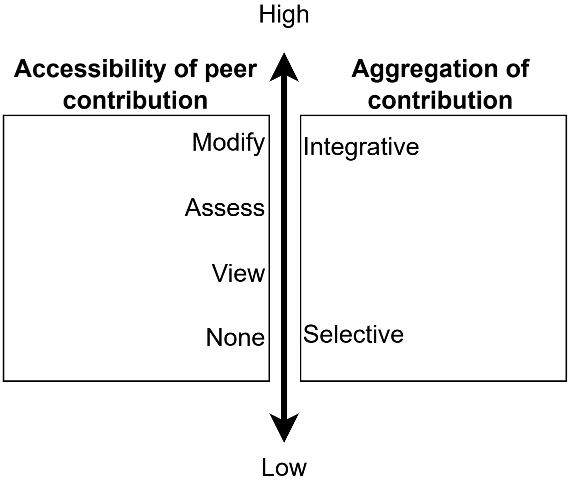


Figure 2.5: Managing crowd contribution concerns the level of accessibility of peer contribution (from high to low: modify, assess, view, and no access rights) and the way of aggregating contributions (integrative and selective) [188].

Unpaid recruitment strategies are used for explicit and implicit contributions. One such strategy is *asking for volunteers*, which relies on crowd participation without monetary compensation, but with rewards such as learning opportunities, social recognition, personal satisfaction, or symbolic rewards (e.g. acknowledgements). Participants in this category are mainly motivated by intrinsic factors and do not expect monetary compensation (e.g., GalaxyZoo). However, recruiting and maintaining voluntary participation can be challenging, as non-paid subjects often represent a minority in the crowdsourcing marketplace. Non-profit organizations often use this model to explore, transcribe, and review digital collections [70, 113].

Another unpaid strategy, *making users pay for service*, integrates crowd participation with services that users receive in exchange for contributing to sub-tasks. For example, users might be asked to solve puzzles (captchas) to access a blog or write articles. These puzzles are part of larger projects such as ReCaptcha, which rely on multiple users for improved detection. Lastly, as mentioned, *piggybacking on user traces* exploits users' behaviour, such as cookies and search engine queries, to recruit participants [145]. The challenge of this approach is determining how to leverage user traces for system purposes. Piggybacking on user traces and making users pay for services represent relatively new forms of Human Intelligent Tasking that are increasingly replacing more standard forms of micro-tasking.

2.2.6. Exploring Dimensions of Workflow Management in Crowdsourcing

In crowdsourcing systems, effective workflow management is essential for ensuring the optimal use of crowd workers' skills, time, and resources. Properly managing the flow of tasks and contributions within the system can significantly enhance overall efficiency, productivity, and outcome quality. This section will explore the critical dimensions of workflow management in crowdsourcing initiatives, including the accessibility of peer contributions and the aggregation of these contributions.

Most crowdsourcing systems, especially those designed for micro-tasks, follow practices that regulate the influx of crowd workers, jobs, and requesters. In the realm of management practices in crowdsourcing, several dimensions play a crucial role in determining how crowd workers can access others' contributions and how their work is aggregated and remunerated (see Figure 2.5). Understanding these dimensions is vital to designing effective crowdsourcing initiatives and harnessing the power of the crowd [188, 505].

One of these dimensions is the *accessibility of peer contributions*, which determines the level of flexibility and involvement in the project [188]. There are four levels of accessibility: *modify, assess, view,* and *none*. At the highest level, "modify" allows contributors to alter, delete, update, and correct each others' work in highly collaborative and community-oriented settings [24]. This level of accessibility can be challenging to monitor by the service provider centrally, as users are in charge of maintaining and editing content. "Assess" degree involves rating, voting, or assessing others' contributions, expected in user-generated content such as digital stores [121], social media sites, and rating services [94]. "View" allows participants to see others' work without being able to modify or assess it, as in public design contests where viewing is part of the process [517]. Lastly, the "none" degree prevents users from seeing other users' work, isolating contributions and avoiding critique, discussion, or manipulation, which may be preferred by organizations with privacy or diversity concerns [342].

Another critical dimension is the *aggregation of contributions* [493], which refers to how companies intend to use crowd contributions. Two types can be distinguished: *integrative* and *selective* [188, 505]. Integrative aggregation pools complementary input from the crowd to tap into their creative power [588] or collective opinion [595]. In contrast, selective aggregation discerns excellent and bad contributions. It selects only the best ones, as seen in contests such as those launched by Innocentive, where part of the selection process involves voting and popularity.

2.3. Crowdsourcing research methods

This thesis presents a collection of studies on collaborative crowdsourcing. The crowd participates in experiments and studies as part of a team or as individuals, providing insights into crowd-team formation and system design. Crowdsourcing research engages large, diverse groups of people in various research processes, including problem-solving, data collection, analysis, and idea generation [572]. To address concerns such as data quality assurance, bias management, and participant engagement, it is crucial to adhere to well-established methods and protocols that ensure standardization, integrity, and reproducibility in research [475, 5]. Below, we outline the standard practices and properties in crowdsourcing research to promote the generation of reproducible and valid results (summarized in Table 2.4).

Table 2.4: Main research methods, properties, and objectives in crowdsourcing studies [475].

Crowdsourcing Research	**Attributes**	**Objectives**
Mapping conditions to tasks	Random assignment	Testing different conditions
Implementation of study conditions	Simultaneous or sequential execution	Avoid repeating participants Control duration and timing
Target population	Sampling mechanisms	Representative sample Increase validity
The task	Interface Instructions	Increase self-selection and task quality
	Task interface source Time allotted	Reproducible research
	Reward strategy	Participants' motivation and output quality
	Quality control	Minimize erroneous contributions
The outcome	Data processing	Dataset validity
	Requester's study design protocols	Ensure ethical research and fair compensation

Mapping conditions to tasks. In comparative studies involving crowds as human subjects, mapping different conditions to tasks is essential to differentiate between results effectively [53]. For example, in a crowdsourced study comparing user interface (UI) designs, researchers might evaluate the impact of various conditions on task completion time and accuracy. Task-to-worker assignment is typically done through randomization. However, different types of assignments are possible, such as matched-pairs-design (i.e., participants are first matched based on relevant characteristics; then, pairs are randomly assigned to other conditions), stratified random assignment (i.e., participants are first divided into distinct strata based on a specific characteristic and then randomly assigned to conditions within each stratum), and more. By *assigning* participants to tasks (or conditions) and ensuring they perform the same job, researchers can analyze the data to identify significant performance differences. This helps draw meaningful conclusions and provide recommendations for future improvements.

Implementation of study conditions. In study research, researchers manipulate specific variables while keeping others constant to observe the effects on the outcomes. Tasks can be executed *simultaneously* (e.g., several groups are performing a given task at once) or *sequentially* (e.g., the same groups execute a sequence of steps or tasks)[2]. Simultaneous execution speeds up data collection and may provide more varied responses. In contrast, sequential execution allows for more controlled comparisons but might take

[2] Simultaneous and Sequential study conditions typically connect with more general experimental study designs such as within and between subjects.

longer to complete and be susceptible to learning effects or fatigue[3]. Researchers must control for *repeat participants* to ensure valid results, as this issue poses a significant challenge in crowdsourcing studies. Monitoring remote participants can be difficult, and mitigating confounding factors such as task familiarity and experience is essential. Implementing strategies such as unique identifiers, screening questions, or technical measures such as cookies or IP tracking can help address this concern and maintain the integrity of study findings. Defining the *study duration and timing* is crucial for reproducibility and understanding demographic differences among samples, as it influences the participant demographics and allows replication under similar conditions. Identifying *input dataset properties* such as origin and public availability is essential for assessing a study's reproducibility and validity in micro-tasking contexts. Finally, conducting *pilot studies* before the main experiment provides valuable insights into the study design, including the rationale for including or excluding specific features.

Target populations. The target population refers to the group of people from which the participants are selected – which, in the case of crowdsourcing studies, is *the crowd*. Deciding whether and how to *filter*, this population can impact the final results [189, 113]. In some cases, *no eligibility criteria* are applied, and the participants are *pre-screened* by the crowdsourcing system without any additional prerequisites. Other times, researchers may use specific criteria such as demographics, digital interface preferences (e.g., mobile or desktop), the number of tasks completed, or the task acceptance rate (i.e., the proportion of tasks that participants accept compared to the total number of tasks offered).

On the other hand, the *sampling mechanism* determines the diversity of the target sample. Crowdsourcing participants generally offer greater variety than subjects in traditional laboratory settings, enhancing the study design's external validity [4]. However, the work completed in crowdsourcing environments can be more challenging to monitor, potentially leading to non-random attrition and impacting internal validity [181]. Thus, it's crucial to carefully consider the sampling strategy to ensure the reliability and validity of the study's results [477].

The task. The *task* is an essential aspect of any crowdsourcing study, and its various attributes can significantly impact the study's outcome (see Section 2.2.2). Some key characteristics of a crowdsourcing task include *task type, task interface, instructions, task interface source, time allotted*, and *reward strategy* [588, 555, 493]. By defining these attributes, researchers can consider organizational and operational details that affect the study's results.

Task type refers to the nature and goal of the study, which influences the crowd's decision to participate in the task. The task interface and instructions combined determine the user interface and guidance the workers receive during the study. Poorly designed interfaces and unclear instructions can negatively affect task execution, causing confusion and resulting in low-quality data [613, 474, 11, 609].

[3] Only if within a design.

[4] Crowd participants tend to have an affinity for technology since most of their tasks require some familiarity with digital interfaces.

Task interface source refers to how the interface appears and behaves, which is crucial for reproducible research in crowdsourcing studies. The time allotted for the task is essential for assessing quality and influences the pay rate, with more extended studies requiring more time and funding. Lastly, reward strategies (the amount paid per task) affect worker motivation and output quality [221, 486].

Aside from the interfaces, instructions, interface source, time allotted, and reward strategy, other important factors determine the quality of the task, namely the validity, depth, and relevance of the output, the monitoring protocols to avoid outliers and suboptimal contributions driven by purely extrinsic gains (i.e., a desire to make money as quick as possible) and other confounding factors (i.e., time of the day, distractions, task complexity, and familiarity, etc.).

Quality control is vital for evaluating the output in studies with crowd participants, as diverse crowdsourcing populations can differ in skills, motivations, backgrounds, and behaviour. Standard quality control methods include rejection criteria (e.g., discarding responses that are too short, incomplete, or off-topic.), training (e.g., teaching workers to identify specific objects in an image before beginning an image tagging task) [343], in-task checks (e.g., including a question that asks the participant to choose a specific answer option to confirm they are paying attention)[146, 96, 148, 504], gold items configuration (e.g., incorporating a few questions with known answers in a more extensive survey and monitoring the accuracy of participants' responses to these questions), post-task control checks (e.g., examining the reactions for plagiarism or other forms of data manipulation) [146, 223, 374], the number of votes per item (e.g., having multiple workers label the same image and selecting the most common label as the final output) [314], aggregation methods (e.g., through majority voting to aggregate worker responses in a classification task)[126, 604], and dropout prevention mechanisms (e.g., offering performance-based incentives, providing feedback, or structuring the task in a way that encourages continued participation) [221, 303]. By utilizing these quality control techniques, researchers can safeguard the integrity and validity of their study and minimize erroneous contributions [244, 477].

The Outcome. The outcome of study crowdsourcing is focused on the *validity* (external and internal) of the results. Deriving accurate conclusions from the data depends on several factors, such as the number of participants, contributions, the dropout rate, and demographics. The study's integrity and conclusions can be compromised if any of the mentioned factors perform poorly, deviate from expectations, or become unreliable. To improve the outcome of studies, it is essential to exclude potentially malicious participants and discard data that produce noisy results, such as incomplete responses or incorrect formatting. The outcome dimension also encompasses data processing, which involves manipulating, cleaning, and categorizing results and the resulting output dataset.

Furthermore, the requester, or the person (or team) that hires crowd workers to execute tasks and participate in studies, plays a significant role in the study design [404, 474]. Critical aspects of the requester's part include the platform(s) used, implemented features, requester-worker interactions, ethical approvals, informed consent, privacy

Table 2.5: Four categories of teams (independent, interdependent, multidisciplinary, and interdisciplinary) according to their properties (e.g., similar/dissimilar skill sets, cooperation levels, and interdependence) and real-world examples.

Team type	**Properties**	**Examples**
Independent	Team members' performance does not affect that of others. Jobs can be easily separated and do not require extensive cooperation or coordination.	Sales teams
Interdependent	Mutual dependence and cooperation. Integration of diverse aptitudes, knowledge, or perspectives is of the essence.	Software development teams
Multidisciplinary	Multiple disciplines and no cross-overs. Members with specialized skills divide roles between them. Strengths and perspectives of different disciplines are the primary focus.	Healthcare teams
Interdisciplinary	Multiple cross-over disciplines. Several stakeholders (e.g., industry, academia). Members with specialized skills work towards a common goal. High communication, alliance, and mutual respect among team members.	Climate change research teams

and data treatment, and participation awareness. When conducting crowdsourcing studies, it is essential to consider the environment, additional features, and interactions between the requester and the workers. Ensuring ethical approvals, informed consent, and fair compensation [37, 225, 605, 259] are crucial for maintaining the integrity of the study and its results [190].

2.4. On teams

In Section 2.2.1, we discussed the diverse nature of crowdsourcing, which spans from simple, decomposable tasks to collaborative and competitive projects. This section narrows our focus to team-based crowdsourcing projects, primarily emphasizing macro-tasking and complex problem-solving involving multiple crowd collaborators. We start by defining teams, teamwork, and team dynamics (Sections 2.4.1 and 2.4.2).

2.4.1. Defining a Team.

A team is a set of human or non-human individuals who work together to achieve a common goal [515]. Teams differ from groups because they rely on complementary skill sets to accomplish their objectives. They are coordinated entities in which each member's strengths are maximized, weaknesses are minimized, and true potential is expressed [409]. It is important to note that not all teams are optimal, and their dynamics can vary significantly. Teams with different sizes, objectives, backgrounds, and timelines exhibit distinct behaviours. Factors such as role distribution, personality traits, leadership styles, hierarchical structures, communication quality, and interaction frequency contribute to a team's success [307]. Teams are, therefore, complex entities with numerous variables that interact to shape team dynamics.

To better understand the concept of a team, we first distinguish between *interdependent* and *independent teams*, as well as *multidisciplinary* and *interdisciplinary teams* (see Table 2.5). Following, we explore the multi-faceted analysis of Guzzo and Dickson [215], highlighting the critical components of teams in organizational settings.

In independent teams, members work on their tasks separately, with minimal collaboration or interaction with one another. Each individual is responsible for their work, and the team's overall success depends on the sum of its members' contributions. In this type of teamwork, members may have distinct roles and responsibilities, and their tasks might not be directly connected to those of other team members. Independent teams may be more common in settings where jobs can be easily separated and do not require extensive cooperation or coordination. For example, sales teams have members responsible for their region, with minimal overlapping. Similarly, call centre teams have members sharing the same objectives but working independently and, most of the time, do not affect each other's work.

In contrast, interdependent teams feature team members who rely on one another to complete tasks and reach their objectives. The work of each member is closely connected to that of others, and success depends on effective collaboration, communication, and coordination [45]. In interdependent teams, members often have complementary skills or functions; their duties are interdependent. This type of team is more common in settings where the integration of diverse aptitudes, knowledge, or perspectives is of the essence. For these teams, one member's success directly affects others' success. Examples are software development teams where developers, designers, and testers must collaborate closely to deliver a high-quality product. Various types of interdependence are usually associated with the timing and coordination of tasks between team members. Generally, there is a distinction between pooled, sequential, and reciprocal forms of interdependence. Pooled interdependence features separate functions that shape the final result, such as a gymnastics team. Sequential interdependence involves one unit's output becoming the input for the next, as seen in assembly lines. Reciprocal interdependence is cyclical, with outcomes and inputs flowing between departments, as in a software company.

In a multidisciplinary team, members contribute their expertise independently, each working within the boundaries of their discipline. The team members work together to address a common goal or problem, but their contributions remain separate rather than integrated. Strengths and perspectives of different disciplines are the primary focus. Teammates from multidisciplinary teams tackle a problem from multiple angles without necessarily combining their knowledge or methods into a single, unified approach. These teams combine diverse practices and technical backgrounds [261]. Examples are healthcare teams collaborating to diagnose, treat and manage a patient's health condition and educational project teams, with teachers, psychologists, curriculum designers, and administrators teaming to devise educational programs.

In contrast, interdisciplinary teams involve members integrating their knowledge, methods, and perspectives to create a shared understanding or generate new ideas. This approach often leads to innovative solutions that transcend traditional disciplinary boundaries. This type of teamwork requires high communication, alliance, and mutual respect among team members [313]. Interdisciplinarity requires synthesizing and blending different disciplines' expertise to create a new, complete understanding of an issue or subject. Climate change research teams are examples of interdisciplinarity, as they come together to understand the consequences of climate change better and propose effective mitigation and adaptation strategies for policymakers and communities.

2.4.2. Key Factors in Team Dynamics

Guzzo and Dickson [215] research some of the most critical challenges that teams must address to achieve optimal effectiveness[5]. The primary factors include *cohesiveness, team composition, leadership and team performance, motivation and team performance, team goals,* and other related issues (see Table 2.6).

Cohesion: Team cohesion represents the bond or attraction that unites team members and enables them to work together effectively [83]. It is associated with the level of *familiarity* among members and plays a crucial role in team *effectiveness.* Cohesion depends on individual and team factors and typically develops over time [40]. Early studies on teamwork cohesion reveal a positive correlation between cohesiveness and effectiveness [626, 162, 216]. Later research builds on these findings, enhancing our understanding of cohesion in teamwork and its influence. Shin et al. [523] define cohesion as the team quality that emerges when team members develop a *commitment to tasks, team pride,* and *interpersonal attraction.* Improved cohesion encourages team members to *care for and help* each other, accelerating the learning process [129]. In Chapter 5, we will explore how cohesion factors into crowdsourced teams for emergency response.

Team composition: Team composition is another factor influencing team effectiveness that Guzzo and Dickson [215] describe as the collection and nature of team members' attributes. It is frequently mentioned in studies on teamwork. Campion et al. [76] identified 19 variables constituting team composition, grouped into five categories (job design, interdependence, composition, context, and process) that impact productivity, satisfaction, and managerial judgment (see Figure 2.6).

Job Design: Drawing inspiration from motivational job design theories [610], this category includes: i) *self-management* (team members' ability to plan, execute, and monitor their tasks independently), ii) *participation* (how actively individuals engage in decision-making and problem-solving), iii) *task variety* (the range of activities and skills a job requires), iv) *task significance* (the impact and importance of a task on the team's overall goals), and v) *task identity* (the distinctiveness and meaningfulness attributed to tasks). Teams tackling shared tasks face challenges similar to those faced by individual

[5] Guzzo and Dickson [215] defines effectiveness as an indicative attribute of a) team-produced outputs, b) the consequences a team has for its members, and c) the enhancement of a team's capability to perform effectively in the future.

Table 2.6: Outline of the main team factors (cohesion, team composition, leadership, motivation, goals, and other issues), their variables, and how they may affect teamwork (adapted from Guzzo and Dickson [215]).

Team factors	**Variables**	**Effects on Teamwork**
Cohesion		Teammates familiarity Effectiveness Commitment to task Pride Interpersonal attraction (e.g., caring and helping)
Team Composition	Job Design	Responsibility Decision-making Teammates familiarity (task, role, work, and accomplishments)
	Interdependence	Motivation Enhanced teamwork Collaboration Coordination
	Composition (of attributes)	Effectiveness Cooperation Teammates interaction
	Context	Employee satisfaction Managerial evaluations Performance Productivity
	Process	Effectiveness
Leadership	Flat leadership	Decentralization Autonomy Collaboration Open communication Fast decision-making
	Hierarchical leadership	Transparent chain of command Stability Streamlined decision-making
Motivation	Collective	Team potency Performance
	Individual	Productivity
Goals		Performance Effort
Other issues	Communication	Effectiveness Performance
	Feedback	Motivation Open communication Performance

workers. These aspects help ensure that team members assume *responsibility*, have a *voice in decision-making*, stay *aware of each other's tasks*, *understand the importance of their work*, and *identify* with their team's roles and accomplishments.

Interdependence: In the context of team interdependence [76], several primary forms emerge: i) *task interdependence* (the degree to which team members rely on one another), ii) *goal interdependence* (how much team members share common objectives), and iii) *feedback and rewards interdependence* (the extent to which team members collectively receive feedback and rewards). Interdependence is vital for fostering *motivation* and strengthening teamwork, as it encourages *collaboration* and *coordination* among team members.

Composition (of attributes): Team composition refers to the mix of characteristics that impact teamwork [76]. These attributes include i) *heterogeneity* (the diversity of team members in terms of skills, backgrounds, experiences, and perspectives), ii) *flexibility* (the ability of team members to adapt to changing circumstances, roles, and responsibilities), iii) *size* (the number of members within a team), and iv) *performance* (the level of individual and collective expertise). These composition attributes play a crucial role in shaping team *effectiveness and cooperation*, as they directly influence how team members interact and collaborate to achieve shared goals.

Context: In teamwork, understanding the context is essential for promoting effective collaboration and accomplishing desired outcomes. Vital elements of team context [76] include i) *training*, ii) *managerial support*, and iii) *communication/cooperation* between teams. Proper training equips team members with the necessary skills and knowledge to perform their tasks efficiently. Managerial support offers guidance, resources, and motivation, contributing to a team's success. Effective team communication and cooperation encourage sharing ideas, knowledge, and resources, streamlining decision-making and problem-solving processes. These context characteristics are closely tied to *employee satisfaction* and *managerial evaluations*, essential indicators of a team's performance and productivity.

Process: Lastly, Campion et al. [76] pinpoint four critical components of team processes, including: i) *potency* (the collective belief of team members in their ability to successfully achieve goals), ii) *social support* (the emotional and practical assistance provided by team members), iii) *workload sharing* (the equitable distribution of tasks and responsibilities among team members), and iv) *communication/cooperation* between teams (the effective exchange of information, ideas, and resources among different teams). Overall team processes are associated with *team effectiveness*.

Leadership: Leadership is crucial for achieving teamwork effectiveness [215]. High leadership expectations lead to better performance [153], and exceptional team managers exhibit excellent tactical skills, improving individual team member performances [262]. However, turbulent and dominant leadership settings can negatively impact teamwork [220]. There are two primary types of leadership: flat and hierarchical. *Flat leadership* promotes decentralization, autonomy, and collaboration, fostering open communication and faster decision-making, but may confuse roles and accountability. On the other hand, hierarchical leadership follows a top-down approach with a trans-

2

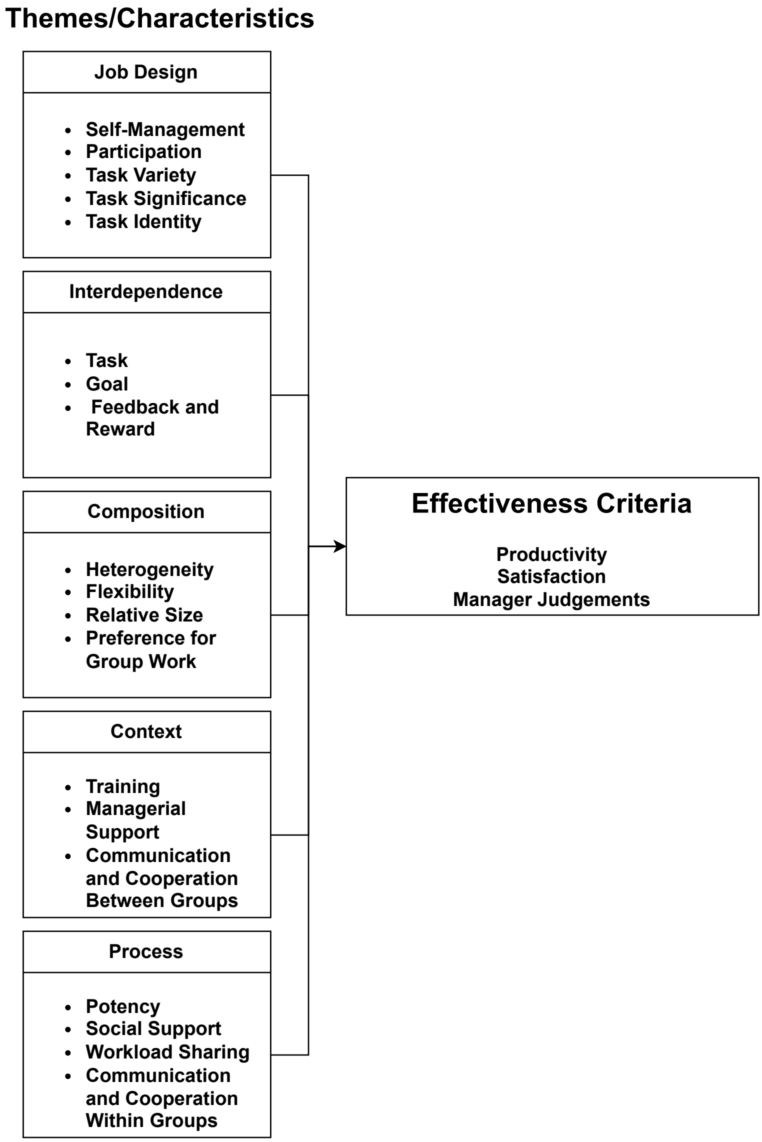


Figure 2.6: Themes and Characteristics Related to Work Team Effectiveness by Campion et al. [76] grouped into i) Job Design, ii) Interdependence, iii) Composition, iv) Context, and v) Process.

parent chain of command, providing stability and streamlined decision-making but potentially limiting communication and innovation. The choice between these styles depends on the organization's size, culture, and goals.

Motivation: Motivation, as a fundamental challenge in team dynamics, significantly influences team *performance* [452] and *productivity* [568]. Guzzo and Dickson [215] distinguishes between *collective* and *individual* motivation. Collective motivation encompasses aspects reliant on multiple team members, such as their estimations, identification, and internalization of roles and responsibilities [521]. Conversely, individual motivation refers to the internal drive for fulfilment and satisfaction, also termed self-motivation. Designing motivational training programs that account for individual differences enhances their efficacy. Collective-focused approaches such as team-building activities are suitable for collectivist individuals. Self-focused methods, such as growth opportunities for team members, benefit individualist people, given their emphasis on team harmony and personal achievement [152].

Team potency denotes the shared belief in a team's effectiveness. Teams exhibiting substantial potency are more likely to attain their goals as they are confident to succeed. Thus, cultivating a sense of potency can effectively motivate teams toward success [215]. Nevertheless, insufficient motivation may lead to social loafing and free-riding in teamwork, although other factors such as dominant personalities could also contribute to such behaviours [435].

Team goals: Establishing clear and well-defined goals is vital for teamwork effectiveness. Guzzo and Dickson [215] emphasizes that goals can address various aspects, including *quantity, speed, accuracy*, or *service to others.* Team goals should indicate what to accomplish rather than the specific means to achieve it. The literature shows the importance of goal setting in teamwork. Weingart and Weldon [598] showed that team goals enhance member effort and performance. Similarly, O'Leary-Kelly et al. [432] found that teams with defined goals significantly outperform those with low goal setting. Höpfner and Keith [248] evaluated Locke's theory of goal setting [346] and confirmed its practical viability for boosting employee motivation and performance.

Other issues: Additional factors influencing teamwork effectiveness include *communication and feedback*, as noted by Guzzo and Dickson [215]. Recent research has investigated the impact of communication styles on team performance. Den Otter and Emmitt [134] highlighted that successful teams employ synchronous and asynchronous communication tools. Effective communication tools, proper training, and robust management competencies are crucial for team effectiveness. Marlow et al. [373] found that communication quality has a stronger association with team performance than communication frequency and that regular face-to-face interactions enhance communication and performance.

Pearson [446]'s research detected modest yet statistically significant productivity increases due to performance feedback. However, other studies [386] did not observe such effects, although they identified shifts in dominant behaviour among individuals receiving goal-referenced feedback. Well-delivered team feedback can bolster *motivation*,

2

Table 2.7: Most common forms of crowdsourced teams and their strengths and weaknesses.

Crowdsourced team type	**Strenghts**	**Weaknesses**
Innovation	Diverse backgrounds, expertise, and creativity Collaborative environment Flexible team structure	Balancing expertise and diversity Maintaining open communication Lack of clear roles and responsibilities
Learning	Focus on knowledge dissimination Ability to expand and reach wider audiences Collaborative learning and skill development	Ensuring accuracy and relevance Maintaining engagement and participantion Informed overload or miscommunication
Emergency Response	Rapid coordination and action Mobilizing large numbers of voluneers Adaptability to evolve circumstances	Efficient communication under pressure Ensuring safety and organization Identifying and allocating resources

facilitate open communication, and *improve performance* by helping team members identify strengths and areas for improvement [133]. Conversely, poorly delivered feedback may cause confusion, lower morale, and impede performance. Effective team feedback hinges on clarity, relevance, timely delivery, and a supportive and constructive approach [307].

2.5. Crowdsourced Teams

Crowdsourced teams are project-based units that exist until task completion, formed by third parties such as companies or self-assembled by the crowd. These teams often comprise dispersed and diverse members, including solely crowd participants or hybrid teams collaborating with outsourcing company members. Key elements such as communication, goal-setting, motivation, team composition, and cohesion are essential for crowd teams and can impact their effectiveness [215]. Crowdsourced teamwork typically encompasses teams focused on innovation, learning, and emergency response, each encountering strengths and weaknesses in communication, goal-setting, motivation, team composition, and cohesion (see Table 2.7).

Crowd teams for innovation address complex challenges by functioning as Collaborative Innovation Networks (COIN) [197]. They consist of self-motivated individuals with diverse backgrounds and a shared vision collaborating to create novel solutions. Members engage in transparent knowledge sharing, creative collaboration, and social networking. Innovation propels these teams, which often operate with flat organizational structures and decentralization of authority, fostering swarm creativity and collective intelligence. However, these teams often struggle to maintain open communication and balance expertise and diversity. Uncoordinated innovation may inadvertently create information silos or generate overwhelming data, hindering effective communication among team members. Striking the right balance between specialized knowledge and diversity can be difficult, as the team must ensure that the diverse skill sets of its members complement each other and contribute to the overall innovation process.

Crowd teams for learning comprise individuals with shared interests and knowledge [272]; these teams focus on learning (and education) and constitute Collaborative Knowledge Networks (CKN) [197]. These teams facilitate knowledge sharing, creation, and dissemination among members, often extending to larger audiences. They may grow larger than innovation-focused teams, allowing for a broader range of perspectives and expertise to be included. These teams frequently face challenges ensuring accuracy, relevance, and engagement in learning activities. Curating accurate and relevant content becomes critical with the vast amounts of information exchanged online and the many platforms offering educational content. This requires continuous monitoring, feedback mechanisms, and quality control processes to maintain the integrity of the shared knowledge. Maintaining engagement in learning activities can also be challenging, particularly in large teams with diverse backgrounds and skill levels. Learning-focused crowd teams research often explores adaptive learning strategies [281], personalized content delivery [33], and gamification techniques to heighten engagement and effectiveness [411].

Crowd teams for emergency response encompass individuals who contribute to emergency response efforts, such as disaster relief or crisis management [30]. These teams can be centrally coordinated or self-assembled during rescue and emergency operations and tend to operate in real time for information exchange and resource allocation. Emergency response crowd teams are usually in charge of geospatial information, social media monitoring, and crisis mapping to swiftly adapt to dynamic situations. These teams form to mobilize resources, expertise, and volunteers rapidly. However, emergency response teams may encounter challenges in communication, safety, and decision-making under pressure. The time-sensitive nature of emergencies demands efficient and effective communication among team members, often across diverse geographic locations, infrastructures, and technologies. Ensuring the safety of team members and affected populations is paramount, which requires the implementation of robust safety protocols, continuous risk assessment, and clear lines of responsibility.

Information overload, uncertainty, and rapidly changing circumstances can further complicate decision-making in high-pressure environments. To address these challenges, research on emergency response crowd teams typically investigates ways to increase the robustness of systems and protocols, including real-time communication channels, data visualization tools, and decision support systems that facilitate informed and timely decision-making. Additionally, research of this type focuses on training and capacity-building efforts to improve team members' situational awareness, decision-making skills, and safety procedures.

2.6. Limits and challenges of crowdsourcing

Imagine being an individual looking for opportunities online to earn income. You come across various digital marketplaces that promise flexible work schedules, diverse tasks, and a chance to make money on your terms. These platforms aid requesters with tasks such as data collection, parsing, or product evaluation in exchange for flexible working hours, minimal experience requirements, and the opportunity to build credibility as you complete tasks. Additionally, you can take on as many jobs as you like. The marketing is alluring, with numerous successful websites celebrating the triumphs of crowdsourcing, showcasing how businesses can solve problems through countless novel and geographically dispersed contributions. The appeal is hard to resist – crowd work appears to be an ideal way to make a living while being ubiquitous, time-efficient, and free from the constraints of traditional outsourcing companies.

As you begin working on these platforms, however, you quickly realize crowd workers face many challenges and limitations. Long hours, inconsistent income, and the lack of worker protections become apparent as you navigate through the world of crowdsourcing. It becomes increasingly difficult to maintain a work-life balance or ensure financial stability. Furthermore, as a crowd worker, you lack targeted support for diversity and inclusion, and team formation and teamwork can feel unstructured and disorganized. Lastly, you notice that your stress levels are higher as a gig economy worker. The constant need to find your next gig or adapt to changes in your current one can be anxiety-inducing. There's less job security in this line of work, which can lead to concerns about sudden changes in your employment status or income. Additionally, being removed from other employees can make communicating and resolving issues with your projects challenging.

On the other hand, consider the experience of a requester using the same platforms to find solutions to their problems. They can quickly post tasks on multiple competitive crowdsourcing marketplaces, set task duration and payment, and accept or reject workers' output without adhering to traditional hiring procedures. Requesters can filter the participant pool by skill level or demographics, tailoring the workforce to their needs. The convenience and flexibility of the crowdsourcing model enable them to access a diverse talent pool and obtain results within a reasonable time frame and budget.

The stark contrast between the experiences of crowd workers and requesters highlights the power imbalance in the crowdsourcing ecosystem. Requesters enjoy numerous advantages, while crowd workers often deal with sub-optimal working conditions and limited support. Studies have highlighted these challenges and suggested ways to improve the experience for crowd workers [299, 331, 357, 136, 27]. To create effective systems for crowd collaboration, it is crucial to understand crowd dynamics at individual and team levels. This thesis contributes to observing the crowd, understanding its approach to team formation, and addressing cooperative challenges.

Additionally, this thesis emphasizes the importance of incorporating the human factor when designing intelligent systems and AI models, following a User-Centered approach. User-centred design, a crucial aspect of Human-Computer Interaction, ensures that users' needs and interests are met while focusing on a product's usability. By extending these principles to crowdsourcing, this thesis aims to investigate how collaborative online systems can better engage users in crowd team formation and teamwork, ultimately bridging the gap between crowd workers and requesters, enhancing safety [75], customer satisfaction [368], and cost efficiency [179].

2.7. Applying UCD principles to Crowdsourcing

User-centred design (UCD) has been increasingly applied to crowdsourcing systems in recent years. It ensures that these platforms are user-friendly, efficient, and effective for contributors and solution seekers. UCD's application to crowdsourcing systems dates back to the early 2000s with the emergence of platforms such as Amazon Mechanical Turk and InnoCentive. These platforms showcased the potential of harnessing collective intelligence and skills from diverse individuals, emphasizing the need for user-centred experiences that promote participation, collaboration, and efficient problem-solving [587, 3]. As crowdsourcing systems evolved, the application of UCD principles adapted to address growing task complexity and participant numbers. Designers now employ various UCD methods, including user research, usability testing, and iterative design, to develop intuitive interfaces, optimize workflows, and enable effective communication among users. The rise of AI-driven crowdsourcing and machine learning integration has further shown UCD's importance. Achieving a balance between human and machine intelligence in these systems demands a deep understanding of users' needs and preferences, making UCD an essential component of their design and development.

This thesis examines various aspects of crowd teams relevant to societal challenges. By observing crowd teams' needs and behaviours when selecting teammates and collaborating online, we can use this knowledge to develop guidelines for user-centred crowdsourcing systems. For crowd innovation generation, team formation, and online collaboration, we adopt five UCD principles from Norman [423]'s book *"The Design of Everyday Things"* 1. Use knowledge in the world and knowledge in the head; 2. Simplify the structure of tasks, 3. Make things visible, 4. Get the mappings right, and 5. Exploit the power of constraints.

2.7.1. Use knowledge in the world and knowledge in the head

This principle emphasizes constructing conceptual models that provide easily understood guidelines. *"Knowledge in the head"* pertains to the information stored in our memory, such as facts and rules. In contrast, *"knowledge in the world"* involves external information, such as written or visual aids, including signs and instructions. Both types of knowledge are crucial for problem-solving and task completion. These principles also connect to other UCD principles: *learnability* and *memorability* [420]. Learnability refers to how easily users can understand and acquire the necessary skills to use a product effectively. Designers aim for high learnability by creating intuitive interfaces and clear instructions informed by user research and usability testing [445]. Memorability concerns users' ability to recall how to use a product after a period of non-use. High memorability is achieved using consistent design patterns, providing feedback, and structuring information to aid retention [3]. System developers may also incorporate cues and reminders [420]. In this thesis, we consider multiple facets of crowd teams that are highly relevant to present societal challenges (Chapters 3, 4, 5, and 6). We also use observational methods to determine what crowd teams need and do when choosing teammates and collaborating online. Collecting and analyzing the crowd's interactions with the systems and the system users, we elicit novel knowledge of the world and knowledge in the head of the crowd users. Afterwards, we implement the knowledge gained from the findings into guidelines (or manuals) for user-centred crowdsourcing systems focused on user-to-user interaction and user preferences.

2.7.2. Simplify the structure of tasks

This principle focuses on minimizing the cognitive load on short- and long-term memory, ensuring that available actions at any given moment are intuitive, visible, and easy to comprehend [541]. Simplification can be achieved by maintaining primary tasks while introducing new supportive infrastructure or modifying the primary tasks themselves [359]. To accomplish this, designers should gain insights into users' experiences through usability testing and various forms of user research. The goal is to develop more effective and user-friendly products by understanding users' experiences and incorporating findings from usability testing and other research methods [3]. In Chapter 5, we propose using strictly cooperative games to assess the capabilities of crowd teams to deal with stress, time-bound objectives, and interdependence. Given that the task is intended for an ad-hoc, short-lived crowd intervention in emergency response, we required participants to communicate only via text and for a limited time. We also redesigned the task inspired by the video game "Keep Talking Nobody Explodes" to fit the activity's objectives and structure. Our design streamlines the modalities of the original game while keeping a firm anchor on strictly cooperative interactions between players. We also focused exclusively on one challenge (among several) that would require one user (Lead Expert) to guide and another (Defuser) to interact with the space and reach the objective on time (defuse the bomb in the maze).

2.7.3. Make things visible.

The principle of "making things visible" is related to the concept of the "gulf of execution and evaluation." The gulf of execution refers to the mismatch between a user's intentions and what the system allows them to do. In contrast, the gulf of evaluation refers to the degree of ease with which users can perceive and interpret whether or not their action was successful [293]—making things visible bridges this gap by providing information about the system's state in a form that is easy to receive and interpret and matches how people think [157]. According to Don Norman's principles of Universal Design, making things visible is essential for bridging these gulfs. It involves designing interfaces that provide feedback on users' actions and show how they affect the system [424]. A well-designed product should bridge these gulfs through its features and system image [293]. For example, when searching for teammates online for a crowdsourcing contest, an "actively searching" control should signal that the user is seeking collaborators. Another example of making things visible is representing crowd workers' characteristics in a relevant, privacy-preserving, and valuable way. In Chapter 6, we ask the crowd to form teams of learners via a drag-and-drop tool designed specifically for team formation online. To our knowledge, this is the first time the team formation problem involves crowd decision-making where the combination of features (e.g., the teammates' characteristics) are visualized and rendered explicit at the individual- and team levels.

Our approach merges crowdsourcing with team formation through adaptive and interactive technologies and graphical user interfaces. The design of the team formation tool was also the product of a usability engineering lifecycle where incremental steps improved the prototypes through several usability tests before being deployed for data collection.

2.7.4. Get the mappings right.

Mapping is a user-centred principle that joins the computer display of information with the user's conceptual models. Mapping users' conceptual models typically requires performing *task analysis*. This is the process of observing users performing a task in ways that can be decomposed into smaller sub-tasks or steps. Task analysis helps with understanding the users' needs and their context.

More complex processes are often needed to understand the users' knowledge representation and their internal model of concepts. For example, asking users to list or team concepts helps map groupings and orderings associated with the users' mental models [419]. Card sorting is another mapping technique to elicit the users' mental models since it requires ordering a set of concepts into piles. In our study presented in Chapter 6, we carried a card sorting task remotely with individual crowd workers to elicit their approach toward team formation and concept mapping.

2.7.5. Exploit the power of constraints.

Designing systems user-centred means accounting for several constraints affecting performance and effectiveness, such as users' physical and cognitive abilities, design principles and guidelines, technical limitations, and business objectives. In crowdsourcing settings, designing collaborative systems means considering the crowd's characteristics, the recruiters' needs, and the limits that technology presents when working globally and remotely.

For example, users' cognitive heuristics and interface biases condition users' choices and behaviour online [511]. Exploiting these constraints when designing collaborative crowd systems may help prevent unwanted behaviour, such as discrimination between team members and prejudice. In Chapter 4, we perform a set of studies comparing different user interfaces and digital nudging interventions to evaluate the effects of digital constraints on the diverse choices of the crowd.

2.8. Conclusion

In this chapter, we explored the concept of crowdsourcing, its classification methods, and the definition and characteristics of teams and crowd teams. We also discussed the challenges faced by crowdsourcing systems in contemporary digital marketplaces. Finally, we outlined the application of five user-centred design principles to collaborative crowdsourcing systems research. In the upcoming chapters, we will present various studies examining different aspects of crowd teams and their preferences in team formation. Chapter 3 focuses on crowd workers' opinions to identify preferences for profiling attributes in team formation systems. Chapter 4 compares digital interventions designed to influence diverse teammate choices in open innovation projects. Chapter 5 investigates crowd teams' emergency response under pressure, analyzing traits and cooperation styles affecting performance and overall teamwork. Chapter 6 examines how the crowd performs team formation tasks given profiling attributes. Lastly, Chapter 7 discusses and concludes the work while suggesting directions for future research.

3

Crowd Preferences for Team Member Profiling

3.1. Abstract

The growing popularity of professional online services has led to increased use of crowdsourcing tools to outsource projects and establish remote teams. In these online self-assembly team formation settings, profiling information is essential for crowd workers, enabling them to gather knowledge about others and build their virtual identities. However, research lacks on what profiling attributes to use from the perspective of crowd workers. In this chapter, we present the results of an online survey to evaluate crowd workers' *willingness to see* and *perceived usefulness* of profiling attributes, categorized into *surface-level* and *deep-level* attributes. These profiling attributes were used to quantify group diversity, differentiating between surface-level demographic and deep-level attitudinal characteristics. In team self-assembly systems, 117 crowd participants rated their preferences for profiling attributes about their and other crowd workers' profiles. Crowd workers prefer displaying and viewing surface-level attributes, particularly those related to demographics and social-media features. In contrast, deep-level attributes, including mental states, beliefs, and political affiliations, are less preferred regarding willingness to share and perceived usefulness. Nonetheless, not all deep-level attributes are perceived negatively, as personality, opinions, and values are considered valuable and relevant within crowdsourcing collaborative systems settings.

3.2. Introduction

Online user profiles in crowdsourcing settings are frequently used to connect small crowds and form teams for open collaboration [336] and group experiments [441]. Many crowd collaboration tools are built ad-hoc, typically requiring remote users to populate profiles for team formation. However, these ad-hoc crowdsourcing team formation systems, while creating digital representations of crowd workers, may inadvertently hinder self-disclosure by requesting intrusive or irrelevant information [478].

Consequently, individuals might hesitate to share or feel compelled to replace truthful disclosure of attributes with false representations of themselves. Missing data or unreliable information can decrease accuracy, potentially undermining trustworthiness [271] and reliability in the outsourced project.

In scenarios such as online crowd team formation, where the accuracy and relevance of self-disclosure are equally important for crowd workers' trust and system viability, it is crucial to assess which attributes crowd workers prefer to disclose about themselves and see about others. Our research concentrates on person-to-person systems designed for self-assembled crowd team formation. These systems, like the research-led application 'My Dream Team' [110], enable users to access other users' online profiles and establish contact points for professional or educational purposes. Users can view their own and others' profiles, which consist of various settings and attributes. Some online crowdsourcing team formation tools function as recommender systems, featuring adaptive filtering capabilities based on users' characteristics and interests [110].

This study examines crowd workers' perception of profiling attributes according to their willingness to see and perceived usefulness. Specifically, we investigate crowd workers' profiling information preferences in self-assembly crowdsourcing team formation systems. These are online platforms where users self-disclose and have access to other users' profiling attributes to form remote teams for open collaborations. We use this study to answer the first Research Question of this thesis:

RQ1: Which personal and professional profile attributes do crowd workers prefer to see and show on crowdsourced team formation systems?

This overarching question is further dissected into more detailed inquiries to understand the preferences of crowd workers comprehensively:

1. **RQ1.1: *About themselves*, which personal and professional profile attributes do crowd workers prefer to display on crowdsourced team formation systems?**
    (a) Which *types* of attributes (surface-, deep-level) are crowd workers *willing to display* about themselves?
    (b) Which *types* of attributes (surface-, deep-level) do crowd workers *find useful* to display about themselves?
    (c) Are crowd workers *more willing to display* surface- or deep-level attributes about themselves?
    (d) Do crowd workers find it *more useful to display* surface- or deep-level attributes about themselves?
    (e) Which *individual* attributes are crowd workers *willing to display* about themselves? Which *individual* attributes do crowd workers *find useful to display* about themselves?

2. **RQ1.2: *About others*, which personal and professional profile attributes do crowd workers prefer to see on crowdsourced team formation systems?**

    (a) Which *types* of attributes (surface-, deep-level) are crowd workers *willing to see* about others?

    (b) Which *types* of attributes (surface-, deep-level) do crowd workers *find useful* to see about others?

    (c) Are crowd workers *more willing to see* surface- or deep-level attributes about others?

    (d) Do crowd workers find it *more useful to see* surface- or deep-level attributes about others?

    (e) Which *individual* attributes are crowd workers *willing to see* about others?

    (f) Which *individual* attributes do crowd workers *find useful to see* about others?

We assess commonly used surface and deep-level profiling dimensions regarding users' willingness to see/disclose and their perceived usefulness on online crowd team formation platforms. Additionally, we evaluate the perceived usefulness of these dimensions within the given context. Surface-level and deep-level attributes are identity-based societal categories describing differences in attributes between people in a workgroup [228]. Surface-level attributes are mainly characterized by their physical and overt nature, easily perceived by others, such as age and gender, and are often part of profiling data. Deep-level attributes encompass people's covert attributes, such as beliefs and attitudes. They are typically acquired through first-hand experiences and are less evident than surface-level traits. To address our Research Question, we conducted an online data-driven study with 117 crowd participants distributed across two surveys:

- **Personal attributes survey: This survey examines crowd participants' preferences for personal profiling attributes**. These preferences are evaluated according to crowd participants' willingness to disclose personal information to other users (called *willingness to see*) and their *perceived usefulness* of personal attributes.

- **Other users' attributes survey: It asks questions only regarding other crowd workers' profiling attributes**. It examines crowd participants' willingness to see profiling attributes about other crowd workers (called *willingness to display*) and their *perception of usefulness.*

The rest of the chapter is structured as follows: Section 3.3 deals with related literature on team formation systems and user profiling attributes. Section 3.4 presents the study design's procedure, metrics, and profiling attributes. It also provides an overview of the crowd participants' demographics. Section 3.5 presents the results from the analysis of the surveys, addressing each Research Question, followed by a discussion (Section 3.6) and conclusion (Section 3.9).

3.3. Crowd Teams in Online Work Environments

The transition from face-to-face to online team formation, particularly for distributed, self-built, and self-organized teams, is expected to grow as virtual remote environments become more prevalent in various aspects of life, such as work, learning, and socializing. With millions of users engaging in remote collaboration [118], distributed projects involving employees with limited shared work history are becoming increasingly common. This trend is not only evident in large multinational organizations [243] but also in crowd-working communities [358] and remote education [323].

Online team formation systems, which are professional social networks consisting of individuals with diverse attributes related to workplaces or educational settings [143], have primarily focused on implementing algorithms for top-down team assembly processes [264, 594]. However, these systems generally do not allow users the option to self-assemble. Recently, innovative team formation systems (e.g., My Dream Team [110] and TeamGen [143]) have emerged, enabling users to connect with other remote users and find collaborators actively.

As demand for online team formation tools based on social network models increases, two types of data are crucial for successful virtual team formation: *profiling attributes* and *user relationships* [143]. Profiling attributes provide general information about a user, such as gender, age, name, role, and skills. User relationships refer to social connections among users, including shared research groups or project assignments. Both profiling data types are commonly used in user modelling and recommendation personalization.

This descriptive information enhances users' profiles with attributes visible to others and used by the system. In addition to demographic data (e.g., username, email, country of origin) and work or education-related information (e.g., role, years of experience, projects), most team formation systems do not adequately address the balanced between profiling attributes suitable for users and systems. As systems become increasingly data-driven, users tend to be more privacy-conscious [545].

Through an initial overview of the literature concerning team formation tools, we notice a disparity in system choice attributes. Most online team formation systems described in Table 3.1 adopt user profiling information according to their pertinence within a discrete context. For example, systems designed for work team formation disclose users' roles (TeamGen [143], GitHub [571], Yammer [589]), Linkedin [550]). At the same time, systems designed for team formation of learners display users' grades and disciplines (e.g., CATME [323]). Other attributes such as surface-level demographic information (gender, age) are shared across most of those platforms (e.g., My Dream team [110], SOT [358], Hive [501], CATME [323]) and are therefore most recurrent.

Profiling information can play a significant role in self-assembly processes. In the study by Gómez-Zará et al. [202] on the effects of displaying personal information [202] on the choice of diverse teammates, certain information negatively impacts diversity. Displaying diversity scores as a profiling attribute can deter users from selecting others dissimilar to them, thus exacerbating network segregation [202]. Although these

Table 3.1: Team formation systems observed in the literature with corresponding profiling attributes and related work.

Team formation system	Users' attribute	Reference
My Dream Team	**Demographic information**; **Competence**; **Soft Skills** (creativity, leadership experience, psychological collectivism, social skills, personality); **Bonding capital**; **Bridging capital** (popularity, activity, betweenness, closeness).	[109]
TeamGen	**Profile attributes** (username, role, skills, location); **Relations** (common project assignments).	[143]
SOT	**Demographic information** (race, age, gender, background education, work status); **Writing experience**; **Creativity level**; **Sample story**.	[358]
GitHub	**Username**; **Bio** (work history, projects; interests); **Contributions history** (issues and pull requests, commits, public, private, and anonymized contributions); **Projects** (repositories, activity in organizations, teams); **Badges**; **Status availability**.	[571]
Hive	**Demographic information** (gender, location, age); **Areas of expertise** (experience with disabilities, design experience, programming experience); **Availability** (scheduling conflicts). Note: this system used the collaborative online tool Slack [274] for team formation and work.	[501]
Huddler	**Username**; **Familiarity** (history of collaboration) **Availability** (response time). Note: like Hive, this system relied on Slack [274] for handling most of the team processes.	[502]
CATME	**Demographic information** (name, gender, race/ethnicity); **Study related attributes** (schedule, prerequisite courses, discipline, grade-point average, sub-discipline, leadership preference); **Attitudes** (big-picture/detail-oriented, commitment level, leadership preference); **Skills and hobbies** (writing skills, software skills, hands-on skills, shop skills, sports, fraternity/sorority).	[323]
Yammer	**Expertise**; **Leaderboards** (most messages, replied-to messages, liked messages); **Member directory**; **Org chart** (managers and reports, list of coworkers); **Praise** (accomplishments and badges); **Personal information** (picture, contact details)	[589]
Linkedin	**Username**; **Location and position**; **Company information** (colleagues, projects, vacancies, employment time); **Publications** (articles, patents, certifications); **Education** (courses, degrees); **Languages**; **Skills**; **Hobbies and Volunteering**;	[550]

studies analyze users' behaviour in the presence of profiling attributes on online self-assembly team formation systems, they do not investigate users' explicit preferences for self-disclosed online profiling attributes. In this research, we surveyed 117 crowd participants by asking questions regarding their preferences of user profiling attributes according to their willingness to see (and disclose) and perceived usefulness.

3.3.1. Workplace team diversity's approach to user profiling

In the field of management sciences, user profiling has been extensively explored under team diversity. In light of this paradigm, users' explicit and self-disclosed attributes are distinct from overt characteristics. Consequently, profiling attributes are described as either surface-level or deep-level.

This neat distinction of profiling information has the advantage of applying to most characteristics found in team formation systems. It also applies to self-disclosed, explicit (or static) attributes that are not necessarily behavioural. For this reason, we follow a similar classification for this study when analyzing users' preferences for self-disclosed features online.

- *Surface-level traits* These are attributes that describe: *"Diversity in the form of characteristics of individuals which are readily visible including but not limited to, age, body size, visible disabilities, race or sex"* [304]. Surface-level traits are biographical characteristics that are easily noticeable in a person. They are the most overt features and do not necessarily reflect people's thoughts or beliefs. Surface-level traits trigger most first-hand reactions, including stereotyping behaviours. These are generally age, gender, ethnicity, country of origin, and education[1]. Considering user profiling online, we add to this category other characteristics typical of social networks and visible to others. These are availability, profile photo (appearance), rating, and popularity. In self-assembly team formation systems, these are traits we expect to be more likely revealed by users since they are straightforward to describe and less ambiguous or sensitive than deep-level traits.

- *Deep-level traits* These are attributes that describe: *"Diversity in characteristics that are non-observable such as attitudes, values, and beliefs, such as religion"* [304]. Deep-level traits are characteristics not noticeable right away. They are more challenging to identify at first glance, yet they influence relationships more than surface-level traits. We identify deep-level traits as the collection of interests, preferences, opinions, personalities, beliefs, and mental states. In self-assembly team formation settings, we expect deep-level traits to be less likely disclosed since they represent individuals' hidden and personal characteristics.

[1]Although education is often classified as a surface-level trait [304], it goes beyond what can be observed on the surface. It encompasses a person's knowledge, skills, and abilities, which are more covert.

Table 3.2: Categorization of profiling attributes adapted from Gong [203] with an additional class of attributes (social-media features).

Type	Attributes	Representative Works
Surface-level traits	*Demographic attributes*	**Age** [479, 417, 340], **Gender** [479, 344, 102], **Location** [479, 363, 337], **Ethnicity** [86], **Education** [586, 397]
Deep-level traits	*Interests and preferences*	**Topical interests** [542, 394, 93, 617, 615, 137], **Geo-preferences** [339, 338]
	Opinions	**Political affiliation** [341, 479, 105], **Topics** [184]
	Personalities	**Personality traits** [199, 305, 19, 31], **Personal values** [92]
	Mental states	**Well-being, Mood, Depression** [128, 127]
	Beliefs	**Religion** [418, 586]
Surface-level traits	*Social-media features*	**Availability, Profile photo, Rating, Popularity** [482, 543]

3

3.4. Study Design

3.4.1. Procedure

We designed two surveys concerning users' *willingness to disclose* and *perceived usefulness* of displaying profiling attributes on online team formation systems. The two surveys are the following:

1. **Personal attributes survey:** Users indicate, on a 5-point Likert scale, which of the 19 attributes of the seven classes (Table 3.2) they prefer to display about themselves on online team formation systems. We evaluated participants' preferences according to their willingness to disclose and perceived usefulness (Section 3.4.2).

2. **Other users' attributes survey:** Users answered the same questions as from the 'personal attributes' survey, with the difference that the profiling attributes concerned other users instead of themselves.

3.4.2. Metrics

We used two metrics to evaluate users' preference for profiling attributes, namely *willingness to disclose* and *perceived usefulness.*

1. *Willingness to disclose* deals with the propensity to disclose personal information, particularly on professional online services voluntarily. This metric is used on participants' profiling attributes (willing for others to see) and other users' profiles (willing to see).

2. *Perceived usefulness* deals with the perceived utility of the disclosure of information in terms of practical worth or applicability within the team formation context.

We applied this measure to participants' profiling attributes (perceived usefulness of their profiling information) and other users' profiles (perceived usefulness of other users' profiling attributes).

3.4.3. User profiling attributes

For the profiling attributes, we adopted part of the classification by Gong [203] (Table 3.2). We modified the classification of the attribute religion, which, in our case, belongs to beliefs[2]. The profiling modelling used for the surveys is as follows:

1. *Demographic attributes* (age, gender, location, ethnicity, education)
2. *Interests and preferences* (topical interests, geo-preferences)
3. *Opinions* (political affiliation and opinions on topics)
4. *Personalities* (personality traits and personal values)
5. *Mental states* (well-being, mood, depression)
6. *Beliefs* (religion)

In addition to these attributes, we added a seventh class called *social-media features* comprising of availability, profile photo, rating, and popularity of the user profile (numbers of views, favourites, comments, etc.). Furthermore, following the workplace team diversity approach, we grouped the attributes into surface-level and deep-level traits (Section 3.3.1).

3.4.4. Participants

Participants were recruited from the crowdsourcing platform Prolific [439]. 117 subjects participated in two separate surveys, with 64 participants in the **Personal attributes survey** and 53 in the **Other Users' Attributes Survey**. Of the participants, 97 reported their nationality as being from various European countries, including Portugal, the United Kingdom, and Poland. Additionally, 20 participants came from other continents, including North and South America and South Africa. For more information on the participants, please refer to Table 3.3. We adhered to Prolific's policy on fair compensation[3] and remunerated participants at a rate of 7.50 GBP per hour [468].

3.5. Results

In this section, we present the findings from the analysis of the surveys addressing the Research Questions summarized in Table 3.4. Our data showed a non-normal distribution, which led us to choose non-parametric statistical tests for our analysis[4]. Due to the large number of tests conducted (44 per dataset), we applied the Bonferroni

[2]Gong [203] placed the attribute Religion in the category Demographic attributes.
[3]As of July 2021.
[4]The Shapiro-Wilk test indicated that the distribution of responses in our dataset significantly deviated from a normal distribution (p-values < 0.05).

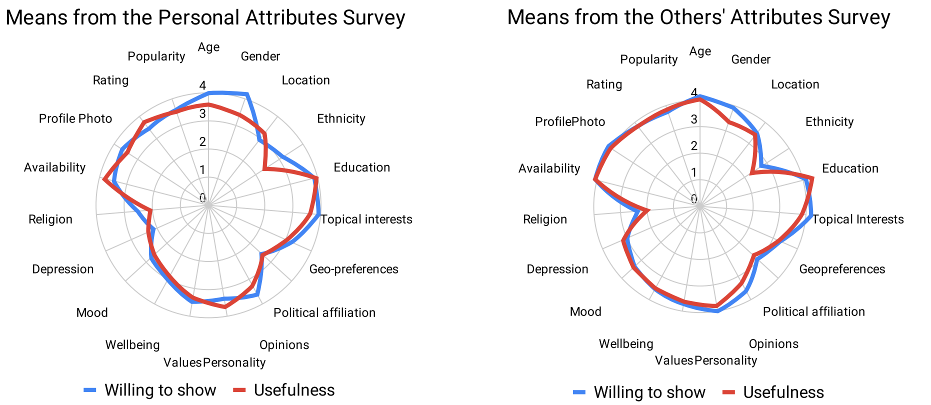
Table 3.3: Demographic information of the participants (N=117) divided by Survey type (**Personal attributes survey** and **Other users' attributes survey**).

Demographic Info		**Personal Attr. Survey**	**Other Users' Attr. Survey**
Nationality	Portugal	21	11
	Poland	8	5
	United Kingdom	7	7
	Greece	4	5
	Spain	2	4
	South Africa	4	3
	Canada	3	4
	Italy	3	3
	Other	12	11
Employment	Job seeking	16	11
	Full-Time	16	20
	Other	15	5
	Part-Time	9	8
	Due to start	2	2
	Not in paid work	2	1
Gender	Female	28	28
	Male	36	25
Student Status	Yes	39	26
	No	20	20
Age	18-24	35	31
	25-34	17	12
	35-44	5	1
	45-54	2	2
	55-64	_	3
	Undisclosed	5	4

3

Table 3.4: Overview of the overarching Research Question paired with the Research Questions and corresponding sub-questions. The table shows the dataset (Survey), number of statistical tests, comparison (Attributes comparison), and metric (willingness to show/see and perceived usefulness) used in the analysis.

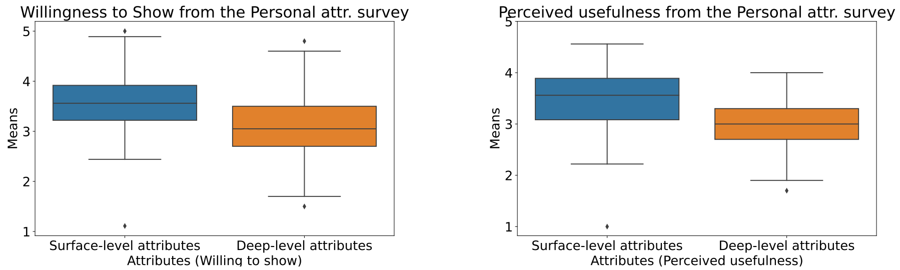
Thesis' RQ		Sub-RQs	Dataset	N. Tests	Attr. comparison	Metric
RQ1	RQ1.1	**RQ1.1.a**	Personal attr. survey	2	Surface-level and deep-level vs. mean (3)	Willing to show
		RQ1.1.b		2	Surface-level and deep-level. vs. mean (3)	Perc. usefulness
		RQ1.1.c		1	Surface-level vs. deep-level	Willing to show
		RQ1.1.d		1	Surface-level vs. deep-level	Perc. usefulness
		RQ1.1.e		19	Each attribute vs. mean (3)	Willing to show
		RQ1.1.f		19	Each attribute vs. mean (3)	Perc. usefulness
	RQ1.2	**RQ1.2.a**	Other users' attr. survey	2	Surface-level and deep-level vs. mean (3)	Willing to see
		RQ1.2.b		2	Surface-level and deep-level vs. mean (3)	Perc. usefulness
		RQ1.2.c		1	Surface-level vs. deep-level	Willing to see
		RQ1.2.d		1	Surface-level vs. deep-level	Perc. usefulness
		RQ1.2.e		19	Each attribute vs. mean (3)	Willing to see
		RQ1.2.f		19	Each attribute vs. mean (3)	Perc. usefulness



(a) Mean of the participants' preference for profiling attributes from the **Personal attributes survey**.

(b) Mean of the participants' preference for profiling attributes from the **Other users' attributes survey**.

Figure 3.1: Participant preferences for profiling attributes from the Personal and Other users' attributes surveys. The means rating for Willing to show/see are in blue, while those for Perceived usefulness are in red.



(a) Means for surface- and deep-level attributes according to the willingness to show.

(b) Means for surface- and deep-level attributes according to the perceived usefulness.

Figure 3.2: Participants' means for surface- and deep-level traits according to the **Personal attributes survey**.

correction to adjust for the risk of false positives[5]. The results are organized and discussed about each specific Research Question. Figures 3.1a and 3.1b provide an overview of the average ratings for personal and other users' attributes according to the participant's willingness to show/see and perceived usefulness.

3.5.1. RQ1.1: *About themselves*, which personal and professional profile attributes do crowd workers prefer to display on crowdsourced team formation systems?

To address RQ1.1, we performed an in-depth analysis exploring a series of sub-questions (i.e., RQ1.1a, RQ1.1b, RQ1.1c, RQ1.1d, RQ1.1e, RQ1.1f), each focusing on personal attributes disclosure on crowdsourcing team formation systems. These aspects include what crowd workers prefer to display and what they find helpful. The analysis is based on data from the **Personal attributes survey**. The significant results adhere to the adjusted p-value=0.001.

RQ1.1.a Which *types* of attributes (surface-, deep-level) are crowd workers *willing to display* about themselves?

Figure 3.2 shows the variation in participants' ratings for surface and deep-level traits. The data's borderline normality led us to use the Wilcoxon Signed-Rank test, comparing responses to a neutral mean of 3. The test shows a pronounced preference for surface-level attributes ($Z = -6.21$, $p < 0.001$). Conversely, the willingness to display deep-level attributes shows no significance ($Z = -1.65$, $p = 0.36$). This pattern suggests a clear preference: **participants are significantly inclined to display surface-level attributes about themselves**.

[5]The Bonferroni correction was used to account for the multiple tests conducted, resulting in a stricter significance level of 0.001.

RQ1.1.b Which *types* of attributes (surface-, deep-level) do crowd workers *find useful* to display about themselves?

We tested for significant differences in perceived usefulness using a Wilcoxon Signed-Rank test comparing the attribute type with the neutral mean of 3. For surface-level attributes, the results are statistically significant ($Z = -5.14$, $p < 0.001$), indicating a strong perceived usefulness of this attribute type. However, deep-level attributes do not exhibit this trend ($Z = -1.72$, $p = 0.72$). These findings highlight that **participants consider surface-level attributes significantly useful**.

RQ1.1.c Are crowd workers *more willing to display* surface- or deep-level attributes about themselves?

Through a Wilcoxon Signed-Rank Test, we tested whether participants would prefer one type of attribute over the other (i.e., surface- versus deep-level). The test indicates a significant preference for surface-level traits over deep-level ones ($Z = -4.73$, $p < 0.001$), supporting the finding that **participants prefer to show surface-level traits over deep-level ones**.

RQ1.1.d Do crowd workers find it *more useful to display* surface- or deep-level attributes about themselves?

Another Wilcoxon Signed-Rank Test comparing surface- versus deep-level traits reported significant differences in perceived usefulness ($Z = -4.74$, $p < 0.001$), indicating a clear preference for surface-level attributes. This finding aligns with the earlier results, suggesting that **crowd workers find surface-level attributes more useful to display than deep-level**.

RQ1.1.e Which *individual* attributes are crowd workers *willing to display* about themselves?

In our analysis, detailed in Table 3.5 and illustrated in Figure 3.3, we employed the Wilcoxon Signed-Rank Test to assess the willingness of crowd workers to display various individual attributes about themselves. The test was conducted against a neutral mean value of 3, considering each attribute's deviation from this benchmark. Our results indicate a willingness to display specific surface-level demographics and deep-level traits. The surface-level traits Age ($Z = -8.36$), Gender ($Z = -8.56$), and Education ($Z = -8.42$) demonstrated significant deviations from the neutral mean ($p < 0.001$). Social media features such as Availability ($Z = -7.28$), Profile Photo ($Z = -7.83$), Rating ($Z = -7.56$), and Popularity ($Z = -7.14$) also resulted in significant differences from the neutral mean ($p < 0.001$). Regarding deep-level traits, Topical Interests ($Z = -8.94$), Opinions ($Z = -7.42$), Personality ($Z = -7.96$), and Values ($Z = -7.82$) all resulted in significant differences from the neutral mean in willingness to show ($p < 0.001$). Depression ($Z = -6.70$) and Religion ($Z = -7.07$) were regarded significantly negatively ($p < 0.001$), indicating a reserved (negative) attitude towards displaying these deep-level traits.

Conversely, other attributes did not produce significant results. The willingness to display surface-level traits such as Location ($Z = -4.69$, $p = 0.13$) and Ethnicity ($Z = -5.01$, $p = 0.53$) did not yield significant differences from the neutral mean. Deep-level attributes

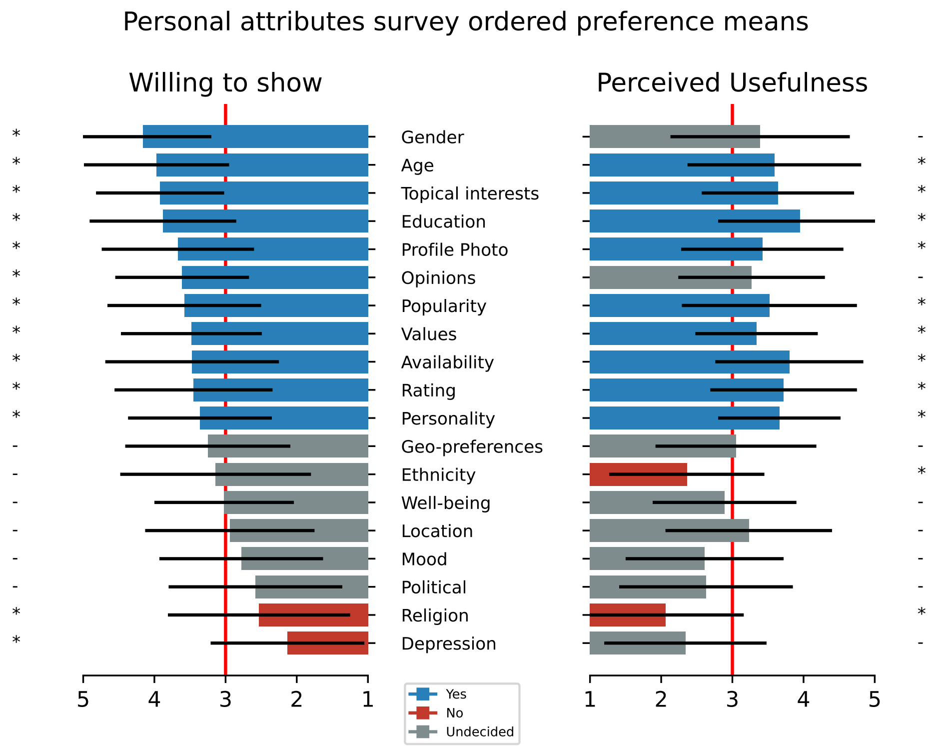


Figure 3.3: Ordered participant preferences for attributes from the **Personal attributes survey**. Bar colours represent whether participants perceived the attributes as significantly positive (Yes), negative (No), or undecided (Undecided) after Bonferroni correction ($p < 0.001$). The lines represent the standard error of the means.

Table 3.5: Mean, Standard Deviation, and significance of the Wilcoxon Signed-Rank Test for all the profiling attributes in the **Personal attributes survey** (N=64). The values are divided by willingness to show and perceived usefulness. The "Show?" column indicates whether the attribute is perceived as significantly positive (y) or negative (n) compared to the scale's mid-point and the adjusted p-value ($p < 0.001$). The last two rows show the average means and std of the means for surface- and deep-level attributes.

			Willing to show					**Perceived Usefulness**				
Type	**Class**	**Attr**	**Mean**	**Std**	**Z**	**p**	**Show?**	**Mean**	**Std**	**Z**	**p**	**Show?**
Surface-level	Demographics	Age	3.97	1.02	-8.36	<0.001	y	3.59	1.22	-7.22	<0.001	y
		Gender	4.16	0.96	-8.56	<0.001	y	3.39	1.26	-5.93	0.00	
		Location	2.94	1.19	-4.69	0.13		3.23	1.17	-4.99	0.02	
		Ethnicity	3.14	1.34	-5.01	0.53		2.36	1.09	-7.92	<0.001	n
		Education	3.88	1.03	-8.42	<0.001	y	3.95	1.15	-8.56	<0.001	y
Deep-level	Interests and preferences	Topical interests	3.92	0.90	-8.94	<0.001	y	3.64	1.07	-7.94	<0.001	y
		Geo-preferences	3.25	1.16	-6.84	0.02		3.05	1.13	-5.83	0.23	
	Opinions	Political affiliation	2.58	1.22	-5.21	0.02		2.63	1.22	-5.65	0.00	
		Opinions	3.61	0.94	-7.42	<0.001	y	3.27	1.03	-6.25	0.01	
	Personalities	Personality	3.36	1.01	-7.96	<0.001	y	3.66	0.86	-8.16	<0.001	y
		Values	3.48	0.99	-7.82	<0.001	y	3.34	0.86	-7.63	<0.001	y
	Mental states	Well-being	3.02	0.98	-6.67	0.02		2.89	1.01	-5.32	0.18	
		Mood	2.78	1.15	-5.00	0.82		2.61	1.11	-5.08	0.57	
		Depression	2.13	1.08	-6.70	<0.001	n	2.34	1.14	-5.42	0.01	
	Belief	Religion	2.53	1.28	-7.07	<0.001	n	2.06	1.10	-8.33	<0.001	n
Surface-level	Social-media features	Availability	3.47	1.22	-7.28	<0.001	y	3.8	1.04	-7.90	<0.001	y
		Profile Photo	3.67	1.07	-7.83	<0.001	y	3.42	1.14	-7.12	<0.001	y
		Rating	3.45	1.11	-7.56	<0.001	y	3.72	1.03	-8.22	<0.001	y
		Popularity	3.58	1.08	-7.14	<0.001	y	3.52	1.23	-6.75	<0.001	y
Surface-level	overall		3.58	0.39				3.44	0.46			
Deep-level	overall		3.07	0.56				2.95	0.54			

such as Geo-preferences (Z = -6.84, p = 0.02), Political affiliation (Z = -5.21, p = 0.02), Well-being (Z = -6.67, p = 0.02), and Mood (Z = -5.00 p = 0.82) did not result significantly different from the neutral mean. These findings highlight that crowd workers show a **selective willingness to display personal attributes, with a clear preference for specific surface-level demographics and deep-level traits, while being more reserved about others**.

RQ1.1.f Which *individual* attributes do crowd workers *find useful* to display about themselves?

Table 3.5 and Figure 3.3 show the results from the Wilcoxon Signed-Rank Test to gauge the perceived usefulness of surface and deep-level attributes against a neutral mean value of 3. The results reveal that surface-level traits such as Age (Z = -7.22) and Education (Z = -8.56) were viewed as significantly more useful than the neutral mean ($p < 0.001$). Other surface-level traits regarding social media features reported significant positive results (Availability (Z = -7.90), Profile photo (Z = -7.12), Rating (Z = -8.22), and Popularity (Z = -6.75), $p < 0.001$). Deep-level traits such as Topical Interests (Z = -7.94), Personality (Z = -8.16), and Values (Z = -7.63) resulted in a significant deviation from the neutral mean ($p < 0.001$), indicating a perceived positive utility for disclosing these traits.

In contrast, the surface-level trait Ethnicity (Z = -7.92) and the deep-level trait Religion (Z = -8.33) resulted in a significantly negative perceived usefulness ($p < 0.001$). Finally, a mix of surface-level (Gender (Z = -5.93, p = 0.00) and Location (Z = -4.99, p = 0.02)) and deep-level attributes (Political affiliation (Z = -5.65, p = 0.00), and Opinions (Z = -6.25, p = 0.01)) did not yield significant results after Bonferroni correction, showing no significant difference from the neutral mean.

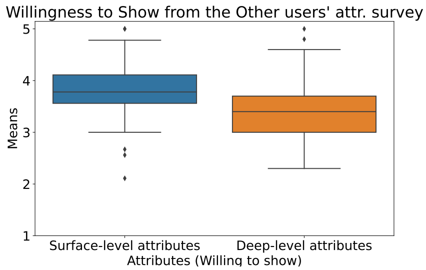
In summary, the results show that **crowd workers perceive surface-level attributes as especially useful to disclose about themselves with deep-level attributes such as personality and values**.

3.5.2. RQ1.2: *About others*, which personal and professional profile attributes do crowd workers prefer to see on crowdsourced team formation systems?

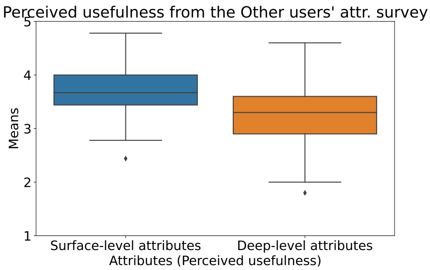
In this section, we extend our analysis to explore crowd workers' preferences regarding the attributes of other users in crowdsourced team formation systems. Similar to the approach taken in RQ1.1, we systematically investigate a series of sub-questions (i.e., RQ1.2a, RQ1.2b, RQ1.2c, RQ1.2d, RQ1.2e, RQ1.2f). Each sub-question examines a different aspect of personal and professional attributes. Still, this time, it focuses on what crowd workers prefer to see about others or find useful in the context of these systems. This analysis uses data from the **Other users' attributes survey**.

RQ1.2.a Which *types* of attributes (surface-, deep-level) are crowd workers *willing to see* about others?

Figure 3.4 presents participants' willingness to see surface-level and deep-level attributes about others. Considering the mixed normality of our data, we used the Wilcoxon



(a) Means for surface- and deep-level attributes according to the willingness to show.



(b) Means for surface- and deep-level attributes according to the perceived usefulness.

Figure 3.4: Participants' means for surface- and deep-level traits according to the Other users' attributes survey

Signed-Rank test, comparing the responses to a neutral benchmark of 3. The results reveal a pronounced preference for surface-level attributes ($Z = -4.21$, $p < 0.001$). The willingness to observe deep-level attributes ($Z = -3.94$, $p < 0.001$) also shows a significant deviation from the mean. This indicates that **crowd workers prefer seeing both surface- and deep-level attributes of others**.

RQ1.2.b Which *types* of attributes (surface-, deep-level) do crowd workers *find useful* to see about others?

Using a Wilcoxon Signed-Rank test, our analysis assessed the perceived usefulness of viewing surface-level attributes compared to a neutral average of 3. The findings highlight a significant difference in perceived utility for surface-level attributes ($Z = -4.97$, $p < 0.001$). Similarly, deep-level traits also exhibit a significant difference from the mean ($Z = -2.53$, $p = 0.011$). However, this is no longer significant after the Bonferroni correction. This pattern suggests that **crowd workers perceive surface-level traits of others as significantly useful** in crowdsourcing team formation settings.

RQ1.2.c Are crowd workers *more willing to see* surface- or deep-level attributes about others?

In addressing this question, we employed the Wilcoxon Signed-Rank test to compare the willingness of crowd workers to see surface-level versus deep-level attributes about others. The test results revealed a significant difference ($Z = -4.97$, $p < 0.001$), indicating that crowd workers **are more willing to see surface-level attributes about other crowd workers over deep-level ones**.

RQ1.2.d Do crowd workers find it *more useful to see* surface- or deep-level attributes about others?

A Wilcoxon Signed-Rank test comparing the perceived usefulness of surface-level versus deep-level attributes revealed a significant difference ($Z = -4.97$, $p < 0.001$). Thus, the results suggest that crowd workers **find it more useful to see surface-level attributes of other crowd workers than deep-level**.

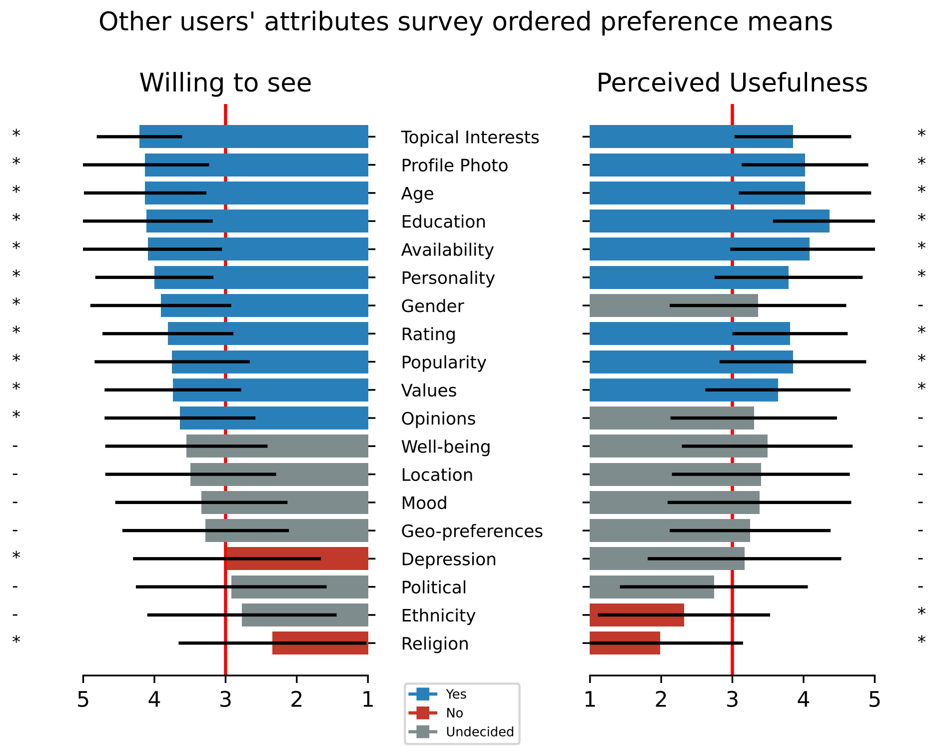


Figure 3.5: Ordered participant preferences for attributes from the **Other users' attributes survey**. Bar colours represent whether participants perceived the attributes as significantly positive (Yes), negative (No), or undecided (Undecided) after Bonferroni correction ($p < 0.001$). The lines represent the standard error of the means.

Table 3.6: Mean, Standard Deviation, p-value and significance of the Wilcoxon Signed-Rank tests for all the profiling attributes in the **Other users' attributes survey** (N=53). The values are divided by willingness to see and perceived usefulness. The "Show?" column indicates whether the attribute is perceived as significantly positive (y) or negative (n) compared to the scale's mid-point and the adjusted p-value ($p < 0.001$). The last two rows show the average means and std of the means for surface- and deep-level attributes.

			Willing to see					**Perceived Usefulness**				
Type	**Class**	**Attr**	**Mean**	**Std**	**Z**	**p**	**Show?**	**Mean**	**Std**	**Z**	**p**	**Show?**
Surface-level	Demographics	Age	4.13	0.86	-8.36	<0.001	y	4.02	0.93	-7.22	<0.001	y
		Gender	3.91	0.99	-8.56	<0.001	y	3.36	1.24	-5.93	0.003	
		Location	3.49	1.20	-4.69	0.128		3.4	1.25	-4.99	0.015	
		Ethnicity	2.77	1.33	-5.01	0.530		2.32	1.21	-7.92	<0.001	n
		Education	4.11	0.93	-8.42	<0.001	y	4.36	0.79	-8.56	<0.001	y
Deep-level	Interests and preferences	Topical Interests	4.21	0.60	-8.94	<0.001	y	3.85	0.82	-7.94	<0.001	y
		Geo-preferences	3.28	1.17	-6.84	0.019		3.25	1.13	-5.83	0.227	
	Opinions	Political affiliation	2.92	1.34	-5.21	0.022		2.74	1.32	-5.65	0.003	
		Opinions	3.64	1.06	-7.42	<0.001	y	3.3	1.17	-6.25	0.012	
	Personalities	Personality	4.00	0.83	-7.96	<0.001	y	3.79	1.04	-8.16	<0.001	y
		Values	3.74	0.96	-7.82	<0.001	y	3.64	1.02	-7.63	<0.001	y
	Mental states	Well-being	3.55	1.14	-6.67	0.019		3.49	1.20	-5.32	0.175	
		Mood	3.34	1.21	-5.00	0.822		3.38	1.29	-5.08	0.570	
		Depression	2.98	1.32	-6.70	<0.001	n	3.17	1.36	-5.42	0.014	
	Beliefs	Religion	2.34	1.32	-7.07	<0.001	n	1.98	1.17	-8.33	<0.001	n
Surface-level	Social-media features	Availability	4.09	1.04	-7.28	<0.001	y	4.08	1.11	-7.90	<0.001	y
		Profile Photo	4.13	0.90	-7.83	<0.001	y	4.02	0.89	-7.12	<0.001	y
		Rating	3.81	0.92	-7.56	<0.001	y	3.81	0.81	-8.22	<0.001	y
		Popularity	3.75	1.09	-7.14	<0.001	y	3.85	1.03	-6.75	<0.001	y
Surface-level	overall		3.80	0.44				3.69	0.61			
Deep-level	overall		3.40	0.55				3.26	0.55			

RQ1.2.e Which *individual* attributes are crowd workers *willing to see* about others?

Table 3.6 and Figure 3.5 show the results from a series of Wilcoxon Signed-Rank tests aimed to understand how the willingness to see other people's attributes deviate from a neutral mean of 3. The findings present a diverse range of preferences. For surface-level demographics, attributes like Age ($Z = -8.36$), Gender ($Z = -8.56$) and Education ($Z = -8.42$) significantly exceeded the neutral mean ($p < 0.001$), showing a strong willingness to see these attributes about others. Additionally, the analysis revealed a significant preference to see surface-level attributes such as Availability ($Z = -7.28$), Profile Photo ($Z = -7.83$), Rating ($Z = -7.56$), and Popularity ($Z = -7.14$) ($p < 0.001$). Deep-level attributes such as Topical Interests ($Z = -8.94$), Opinions ($Z = -7.42$), Personality ($Z = -7.96$), and Values ($Z = -7.82$) resulted in a significant positive perceived willingness to see ($p < 0.001$). However, other deep-level attributes, such as Depression ($Z = -6.70$) and Religion ($Z = -7.07$), resulted in a significant negative willingness to see. Lastly, surface-level attributes such as Location ($Z = -4.69$, $p = 0.128$) and Ethnicity ($Z = -5.01$, $p = 0.530$), as well as deep-level attributes such as Geo-preferences ($Z = -6.84$, $p = 0.019$), Political affiliation ($Z = -5.21$, $p = 0.022$), Well-being ($Z = -6.67$, $p = 0.019$), and Mood ($Z = -5.00$, $p = 0.822$) did not significantly differ from the neutral mean after Bonferroni correction. These results suggest that **crowd workers demonstrate selective preferences in viewing the personal attributes of others, showing an inclination for certain surface-level and deep-level traits while being comparatively neutral or less interested in others**.

RQ1.2.f Which *individual* attributes do crowd workers *find useful* to see about others?

We employed the Wilcoxon Signed-Rank Test to gauge significant differences in the perceived usefulness of attributes from other crowd workers' profiles. This test evaluated the deviation of each attribute's ratings from a neutral mean of 3. For surface-level demographic attributes, traits such as Age ($Z = -7.22$) and Education ($Z = -8.56$) were viewed as significantly more useful than the neutral mean ($p < 0.001$). Similarly, surface-level attributes such as Availability ($Z = -7.90$), Profile Photo ($Z = -7.12$), Rating ($Z = -8.22$), and Popularity ($Z = -6.75$) were seen as highly useful to view ($p < 0.001$). Deep-level attributes such as Topical interests ($Z = -7.94$), Personality ($Z = -8.16$), and Values ($Z = -7.63$) resulted in significantly positive perceived usefulness ($p < 0.001$). In contrast, the surface-level attribute of Ethnicity ($Z = -7.92$) and the deep-level attribute of Religion ($Z = -8.33$) resulted in a significantly negative perceived usefulness ($p < 0.001$).

Finally, other surface-level traits such as Gender ($Z = -5.93$, $p = 0.003$), Location ($Z = -4.99$, $p = 0.015$), and several more deep-level traits such as Geo-preferences ($Z = -5.83$, $p = 0.227$), Political affiliation ($Z = -5.65$, $p = 0.003$), Opinions ($Z = -6.25$, $p = 0.012$), Well-being ($Z = -5.32$, $p = 0.175$), Mood ($Z = -5.08$, $p = 0.570$), and Depression ($Z = -5.42$, $p = 0.014$) did not significantly differ in perceived usefulness from the neutral mean. In summary, these findings suggest that **crowd workers perceive specific attributes, especially surface-level demographics and certain deep-level traits, as more useful to see about others while holding a more balanced view of the usefulness of other attributes**.

3

Table 3.7: Count of attributes resulting significantly positive (Yes), negative (No), or non-significant (Und.) from the Wilcoxon Signed-Rank tests.

		Personal attributes survey						**Other users' attributes survey**								
		Willing to show			**Perceived use.**			**Willing to see**			**Perceived use.**			**Count**		
Type	**Attribute**	**Yes**	**No**	**Und.**	**Yes**	**No**	**Und.**	**Yes**	**No**	**Und.**	**Yes**	**No**	**Und.**	**Yes**	**No**	**Und.**
Surface-level	Age	x			x			x			x			4	-	-
	Gender	x					x	x					x	2	-	2
	Location			x			x			x			x	-	-	4
	Ethnicity			x		x				x		x		-	2	2
	Education	x			x			x			x			4	-	-
	Availability	x			x			x			x			4	-	-
	Profile Photo	x			x			x			x			4	-	-
	Rating	x			x			x			x			4	-	-
	Popularity	x			x			x			x			4	-	-
Total (Surface-level)														26	2	8
Deep-level	Geo-preferences			x			x			x			x	-	-	4
	Political affiliation			x			x			x			x	-	-	4
	Opinions	x					x	x					x	2	-	2
	Personality	x			x			x			x			4	-	-
	Values	x			x			x			x			4	-	-
	Wellbeing			x			x			x			x	-	-	4
	Mood			x			x			x			x	-	-	4
	Depression		x				x		x				x	-	2	2
	Religion		x			x			x			x		-	4	-
Total (Deep-level)														10	6	20

3.6. Discussion

This chapter explored crowd workers' preferences regarding personal and professional profile attributes in crowdsourced team formation systems. Our analysis addressed the main Research Question **RQ1: Which personal and professional profile attributes do crowd workers prefer to see and show on crowdsourced team formation systems?** The findings, derived from our analysis of survey responses, reveal insightful preferences among crowd workers. Table 3.7 presents an overview of how different attributes are rated in our study based on the Wilcoxon Signed-Rank tests. Our analysis categorized attributes into two main groups: surface-level and deep-level. The surface-level attributes, encompassing age, gender, location, ethnicity, education, availability, profile photo, rating, and popularity, showed a predominant acceptance. Specifically, 26 out of 72 instances of all the available attributes (36%) were marked as desirable regarding willingness to show/see and perceived usefulness. Conversely, deep-level attributes, which include geo-preferences, political affiliation, opinions, personality, values, well-being, mood, depression, and religion, were less frequently endorsed, with only ten instances (14%) seen positively. A notable finding is the reluctance to accept specific attributes. Within the surface-level group, ethnicity was mainly resisted for perceived usefulness. This hesitation was more pronounced in the deep-level category of traits like depression and religion.

Interestingly, the study revealed a significant degree of uncertainty among participants. For surface-level traits such as gender, location, and ethnicity, eight instances showed ambivalence. This uncertainty escalated in deep-level traits like geo-preferences, political affiliation, well-being, mood, and depression, with as many as 20 instances reflecting a lack of clear stance. In summary, our results illustrate a clear preference for surface-level attributes over deep-level ones, highlighting a dichotomy between crowd workers' privacy concerns and disclosure preferences in online collaborative spaces. This trend aligns with findings from [13, 277], who also noted similar patterns in online privacy behaviour. Table 3.7 offers a granular view of these preferences, reflecting the nuanced landscape of digital privacy and user engagement.

Our results from the **Personal attributes survey** highlight an apparent inclination towards revealing surface-level attributes. This trend is evident in the preference for showcasing attributes like age, gender, and education. Interestingly, there is a marked reluctance to display deep-level traits, particularly those related to mental states like mood and depression. However, **personality, opinions, and values were regarded more favourably**. When evaluating the attributes of others in these systems, our findings mirror the trends observed previously (i.e., whether and what crowd workers are willing to disclose and find useful about themselves). Participants strongly preferred viewing surface-level attributes such as age, gender, and education of other users. Again, deep-level attributes were less favoured, indicating a consistent pattern in the perception and valuation of personal information, both in willingness to see and perceived usefulness. Nonetheless, participants considered knowing the **personality and values** of other crowd workers valuable and relevant to collaborative crowdsourcing team formation systems.

The combined findings from the **Personal Attributes Survey** and the **Other Users' Attributes Survey** reveal a preference for surface-level attributes among users, with a notable exception for personality traits and values. These exceptions are likely due to their perceived importance in team dynamics and compatibility, which is crucial for our research in developing crowdsourcing systems tailored to user preferences and views. Extensive research, including studies by Lykourentzou et al. [354, 352] highlight the significance of personality and values in the effectiveness of crowdsourcing collaboration teams. Teams with compatible personalities tend to perform better, highlighting the importance of including personality and values in forming crowdsourcing teams. This approach aligns with user preferences, supporting the creation of more user-centred collaborative environments.

Overall, the pattern emerging from our study demonstrates a strategic approach adopted by crowd workers in the context of self-representation and evaluating others in crowdsourced environments. The selective preference for surface-level attributes can be attributed to various factors, including perceived relevance, ease of understanding, and a desire to maintain personal privacy while still engaging effectively in these platforms. Our findings have significant implications for understanding dynamics in team formation and collaboration within crowdsourced environments. The preference for specific attributes over others can influence how teams are formed and individuals interact within these systems [277, 13]. As part of future work, researchers may experiment with displaying surface-level and deep-level traits according to the length and nature of the collaborative task.

3.7. Limitations

While our study provides valuable insights into the preferences of crowd workers in crowdsourced team formation systems, several limitations should be acknowledged. These limitations pertain to the study design, the choice of statistical tests, and other factors that might influence the interpretation and generalizability of the findings.

1. **Study design.** Our study's sample may not represent all demographics equally, potentially limiting the generalizability of the findings. The inherent characteristics of the sample, such as geographic location, age distribution, or professional background, could have influenced the results. Furthermore, relying primarily on survey data may introduce self-reporting and social desirability biases. Respondents might have provided answers they perceived as more acceptable or favourable than their genuine preferences. Another significant limitation is that the study focused on attribute preferences without delving deeply into the reasons behind these preferences. Understanding the motivations and contexts that drive these choices could provide a more nuanced interpretation of the results.

2. **Analysis.** The choice to use non-parametric tests, specifically the Wilcoxon Signed-Rank Test, was based on the non-normal distribution of the data. However, this approach is less powerful than parametric methods in certain circumstances, potentially affecting the ability to detect actual effects.

3. **Other factors.** Another limitation is that crowdsourcing platforms and users' behaviours are dynamic and may change over time. The preferences and behaviours observed in this study can evolve as the platforms and their user bases evolve. Technological advancements and cultural shifts can influence how people interact with online platforms and disclose information. Future studies need to account for these evolving dynamics.

3.8. Implications and Future Directions

This research has highlighted the need for privacy in crowd team formation settings (i.e., crowd workers are reluctant to show deep-level traits) and the need for relevance and utility of disclosure of covert traits such as values, personality, and opinions. The apparent emphasis on personality traits in our findings offers important implications for the design of crowdsourced platforms. It suggests a need to prioritize features that allow users to effectively convey their personality, which might enhance individual profiles and aid in forming more compatible and effective teams. This is particularly relevant for follow-up studies including the one presented in Chapter 5. Comparative studies across different platforms could provide insights into whether these preferences are universal or context-specific. In conclusion, our study highlights the significant role of personality traits in crowdsourced environments, both in terms of willingness to display and perceived usefulness. These findings contribute to the broader discourse on online identity construction and team formation dynamics, offering valuable directions for future research and platform design.

3.9. Conclusion

This study investigated crowd workers' preferences for profiling attributes in self-assembly team formation systems. It addressed the first Research Question of this thesis, namely **"RQ1: Which personal and professional profile attributes do crowd workers prefer to see and show on crowdsourced team formation systems?"** . The study revealed a pronounced preference for displaying and viewing surface-level attributes such as age, gender, education and social media features such as availability and profile photo. This finding indicates a general tendency among crowd workers to favour easily observable traits that facilitate immediate understanding and assessment in online interactions. However, the study also highlights the perceived usefulness and relevance of showing and seeing certain deep-level attributes, particularly personality and values. Showing profiling attributes (particularly surface-level ones) has the potential danger of leading to a lack of diversity in team formation. Therefore, the next chapter evaluates digital interventions to nudge crowd workers towards inclusion and diversity in online collaborative workspaces.

4

Digital Nudging to Enhance Crowd Team Diversity

4.1. Abstract

Companies increasingly want to boost team diversity both for reasons of equality and inclusion, and because of its benefits for team performance. The emergence of self-assembling team formation systems, where online users can select their teammates, unfortunately often reduces diversity, as people tend to choose others similar to them. Research is needed on how to influence crowd workers to create more diverse team. In this chapter, we therefore address the Research Question **RQ2: What is the impact of digital nudging techniques on promoting diversity in self-assembled crowd project teams?** and examine whether making users aware of the team's diversity can impact their selections. We tested the effects of two-choice architecture and nudging techniques in a study involving 120 crowd participants working on a crowdsourced innovation project scenario. The first technique displayed explicit personalized Diversity Information in the form of the current team Diversity Score and diversity recommendations. The second technique used diversity Priming in the form of counter-stereotypes and All-Inclusive Multiculturalism. Our results indicate that Priming deterred participants from picking teammates from different regions and that displaying Diversity Information was the only factor that positively enhanced diverse choices. Other factors we also found to predict selection behaviour were the participants' region of origin, gender, teammates' functional backgrounds (i.e., skills, expertise, and professional experience), and their order of appearance. In light of these findings, we suggest that nudging techniques must be cautiously applied to online team formation as the different techniques differ in their ability to evoke diversity among intrinsically diverse crowds and that personalized displaying of Diversity Information seems most promising.

4.2. Introduction

With a growing international outlook to doing business and outsourcing innovation, diversity and inclusiveness have become substantial parts of most companies' assessments and progress reports [416, 254] while pro-diversity managerial practices are also on the rise [618]. These trends are strengthened by recent findings on the positive effect of team diversity on the organizational goal of innovation (e.g., gender [496], nationality [296], and personality [87] diversity).

Still, employers can mistakenly overlook employees' homophilic preferences for collaborators [49] or are subject to gender stereotyping or unconscious biases [90]. Outsourced crowdsourcing teams can also be subject to homophilic biases and stereotypes while self-assembling and self-organizing [254]. Persisting biases can trigger practices responsible for marginalizing contributors from different backgrounds. Yet, team diversity – especially in open collaboration and crowdsourced innovation projects — is often one of the best assets of crowd collaborative labour [144, 522]. Teams heterogeneous in skills, tenure, and geo-location tend to outperform homogeneous ones in complex and creative tasks [367, 144].

Team diversity can take the form of divergence in opinions and thinking, including political orientation. A diverse political compass amongst contributors can benefit the quality of the teamwork output, as shown in the study by Shi et al. [522]. Their work shows that ideologically polarized Wikipedia teams, such as those composed of the most diverse political slants, are substantially more constructive, competitive, and focused than ideologically homogeneous ones [522].

Despite communication-inhibiting factors [144], diversity aids creative and innovative solutions to complex, open-ended problems [144, 522]. Considering several advantages of team diversity within crowd collaboration [367, 144] and the capacity of digital interfaces to connect diverse collaborators, we ask the following: *how can open collaborative tools support the formation of more diverse crowd project teams?*

Interfaces are known to condition users' choices [395]. The very presence of information while making decisions online can prime users to change their behaviour toward an intended outcome. Images and content prime people to build up assumptions and expectations that guide their thought associations. Gómez-Zará et al. show that profiles with high Diversity Scores are less likely to be chosen by university students forming online teams [202]. These findings suggest a need for more practical, real-world research examining how digital nudging strategies — used in digital platforms such as websites and mobile apps- influence user behaviour and decision-making [599]. While these nudging strategies seem logical and promising, their effectiveness has not yet been fully proven through empirical studies [74].

Combining a growing managerial emphasis on organizational diversity with the growth of open collaborations and crowdsourced innovation projects, we identify a gap in the literature regarding interventions designed to safeguard diversity among self-assembled crowd teams. By self-assembled teams, we mean those teams generated through a bottom-up process where actors self-organize [449, 202]. In a scenario where people choose *The best person for the job* [202, 228], we aim to observe to what degree participants made choices based upon surface-level diversity of their teammates (complexion, gender) versus their deep-level traits (skills and level of education).

We present a study on the impact of Priming and Diversity Information (two digital nudging techniques) on the formation of teams for outsourced crowdsourced innovation projects focusing on a creative complex task representative of crowdsourcing open, diverse creativity and design thinking [316]. Our Research Questions are summarized as follows.

RQ2: How do Priming and Diversity Information affect the diversity[1] of crowd users' team member choices? This question examines two distinct nudging techniques — Priming and the display of Diversity Information — and explores their combined effect, thereby creating three Research Questions.

- **RQ2.1: (How) does Priming affect the diversity of the members that crowd users select for their team?**
- **RQ2.2: (How) does displaying Diversity Information (DI) affect the diversity of the members that crowd users select for their teams?**
- **RQ2.3: (How) does the combination of Priming and Diversity Information (DI) (Priming + DI) affect the diversity of team members that crowd users select for their teams?**

Our study recruited 120 crowd participants to autonomously assemble virtual teams comprising two teammates (plus themselves). Participants were randomly assigned to one of the four conditions (control, Diversity Information, Priming, Diversity Information plus Priming). We found that personalized Diversity Information (DI) positively affected heterogeneity, contrasting with the findings from Gómez-Zará et al. [202]. Nonetheless, participants still chose primarily according to homophilic gender preferences and region of origin. In contrast, the type of task (creating a slogan for a coffee company) seemingly drove crowd participants to choose specific functional backgrounds (skills, expertise, and professional experience) over others. This study offers insights into how socio-technical team formation systems can contribute to more diverse teams in open collaboration for crowdsourced innovation projects. It builds upon previous research on digital diversity interventions [202] and aims to shed light on how technology can play a role in attentively stimulating diversity among crowd collaborators. Furthermore, it identifies which digital interventions among Priming techniques and Diversity Information (including recommendations) could adversely affect diverse choices.

The rest of the chapter is as follows: Section 4.3 covers the related work on crowdsourcing innovation projects, digital nudging techniques, and nudging for diversity. Section 4.4 proposes the Research Questions and relevant hypotheses. Section 4.5 presents the study design. Section 4.6 analyses the results and Section 4.7 discusses these along with system design recommendations gathered from the study. Section 4.8 concludes the chapter.

[1] Metrics for diversity will be discussed in Section 4.5.4 and will consider the team's diversity with respect to age, gender, functional background, level of education, and cultural background.

4.3. Related Work

In this section, we present the related work underlying our exploration of digital nudging in the context of fostering diversity within crowd-sourced innovation projects. Our discussion spans the multifaceted nature of diversity in crowd teams (Section 4.3.1) and the complexities of digital nudging techniques (Sections 4.3.2). We identify opportunities and challenges when applying digital nudges to online diversity and inclusive interventions (Section 4.3.3). Lastly, we explain the specific techniques we are investigating, namely Priming and Diversity Information (Sections 4.3.4 and 4.3.5).

4.3.1. Diversity in crowdsourcing teams

In the dynamic realm of collaborative innovation, crowdsourced innovation projects' emergence and increasing prevalence stand out as a significant development [280, 620]. These projects have become instrumental in shaping the dynamics of contemporary crowdsourcing teams [620]. The widespread acceptance and integration of crowdsourced innovation projects across various domains signal a shift in the approach to collaborative online work [365, 558]. This shift emphasises the importance of cultivating diverse, user-centred teams [632], a concept at the heart of this chapter. The increasing demand for crowdsourced innovation projects is deeply rooted in the capacity of crowdsourcing remote teams to draw upon a broad spectrum of perspectives, skills, and experiences [280, 632]. The positive impact of diverse teammates in open innovation is most felt in projects that demand high creativity, problem-solving insight, and innovative thinking [365, 358]. In this context, understanding how crowdsourced innovation projects operate is pivotal to informing, guiding, and delimiting the direction of our research on diversity and inclusion in online team formation. Crowd teams formed through crowdsourced innovation projects share qualities that can be summarized as follows:

1. **Competitive and Collaborative:** Crowd teams in crowdsourced innovation projects engage in competitive and collaborative tasks [443]. Team members often compete to offer the best ideas or solutions and collaborate to achieve a common goal. This duality fosters an environment where creativity and innovation are paramount and where the diverse skills of team members are harnessed to drive project success.

2. **No Size Limit Unless Specified by the Requester:** These teams are typically not bound by a fixed number of participants. The team size is flexible and can expand or contract based on the project's requirements unless the project initiator (requester) specifies a limit. This flexibility allows for a scalable and dynamic team structure, accommodating a range of project sizes and complexities.

3. **No Hierarchical Structure:** Typically, these teams operate without a formal hierarchical structure. There are no predetermined leaders or rigid roles. Instead, leadership and roles may emerge organically based on the skills and contributions of the members [480]. This lack of hierarchy encourages egalitarian participation and can lead to more democratic decision-making processes [625].

4. **Voluntary Ad-hoc Membership with Fluid Boundaries:** Membership in these teams is usually voluntary and based on the interest and availability of the participants. Team composition can change as members join or leave, allowing for a fluid and adaptable team structure [502]. This fluidity supports a dynamic environment where new ideas and perspectives are continuously integrated.

5. **No Predetermined Division of Tasks:** Unlike traditional teams with predefined roles and responsibilities, these crowd teams do not have a predetermined division of tasks. Instead, collaborators within these teams engage in self-coordination, autonomously deciding on their roles and responsibilities. Each member voluntarily chooses which tasks to undertake, leading to a more organic and fluid task allocation process. This self-coordination is fully autonomous, allowing for a flexible and dynamic approach to completing the project's objectives [209].

Through crowdsourced innovation projects, the crowd is responsible for finding collaborators and is expected to generate innovative solutions. Moreover, open collaboration in crowdsourced innovation projects is based on the understanding that complex, open-ended problems benefit from the diverse expertise and skills typically found in a varied crowd. As previously mentioned, Shi et al. [522] discovered a compelling link between the diversity of collaborators and the nature of the content they produce when analysing Wikipedia talk pages. Specifically, they found that articles with higher debate intensity and richer lexical and semantic diversity were predominantly authored by politically polarized groups, indicating a high level of diversity in viewpoints. This diversity in political thought needed a balancing act from contributors with differing viewpoints, particularly in editing contested topics. Such a dynamic, where politically diverse opinions actively engage and balance conflicting viewpoints, is often absent in more homogeneous communities (i.e., where diversity is low and the team members' attributes are similar to one another), like echo chamber platforms, as noted by Sunstein [554].

However, the relationship between diversity and team effectiveness is complex and multifaceted. Not all forms of diversity yield positive outcomes for teams. While deep-level diversity aspects, such as varied skill sets and tenure, enhance creative problem-solving in crowdsourcing environments [144], there are dimensions of collaboration where homogeneity plays a beneficial role. For instance, homophily – similarity among team members in language, geographical proximity, and familiarity – has been observed to facilitate communication and coordination [144]. This advantage of homophily can often be attributed to shared and acquired characteristics, like past collaborations, common language, and customs, which foster a more synergistic environment for communication and coordination within teams. This nuanced understanding illustrates that *while diversity can bring various perspectives and foster rich, debated content, specific homogeneous attributes among team members can also be crucial for efficient communication and rapport.* The receipt, therefore, may lie in balancing these diverse and homogeneous elements depending on the context and tasks to optimize team performance in collaborative environments.

4.3.2. Digital Nudging

Digital nudging, a concept within the broader scope of persuasive computing technologies, uses digital interfaces to guide individuals' behaviours towards desired outcomes. As defined by Sunstein [552], this approach employs subtle methods of suggestion and positive reinforcement. Fundamental techniques in digital nudging, identified by Weinmann et al. [599], include *default options*, *positioning*, *explanations*, and *decoys*. These techniques capture user attention and direct behaviour in specific ways.

- *Default options* are typically used to capitalize on the human inclination to minimize decision-making effort. This is demonstrated by the widespread use of default privacy settings in social media platforms, which users often accept without modification [201, 238].
- *Positioning* leverages the serial-position effect — a psychological phenomenon where individuals most vividly recall the first and last items in a series [165]. This effect is strategically used in digital environments to emphasize particular choices [36].
- *Explanations* provide users with the context and information necessary to navigate complex decision scenarios. This approach has aided users, especially in environments where decision-making is based on intricate algorithms or systems [47].
- *Decoys* subtly alter users' perceptions of choices by introducing less attractive options, making other options seem more appealing through comparative judgment [444].

While these techniques differ in application, they share a unified goal: **to influence user decisions subtly yet effectively without limiting the range of choices available**. Digital nudging, through these varied interventions, alters user behaviour in digital contexts and indicates the growing sophistication and subtlety of persuasive technologies [172]. Modern recommender systems, (e.g., Netflix and Spotify), exemplify digital nudging. These systems, while aiding users in finding relevant items and avoiding choice overload, also align with the organizational goals of increasing sales and user engagement. These effectively nudge users towards decisions that benefit them without restricting their choice space, thus creating significant business worth [35]. A recent taxonomy of digital nudging mechanisms by Jesse and Jannach [269] grouped the main types of nudges identified in digital environments (Figure 4.1). We provide an overview of the most salient aspects of these categories, namely *decision information, decision structure, decision assistance,* and *social decision appeal.*

Decision Information encompasses nudging mechanisms that provide or emphasize specific information to guide decision-making. These mechanisms focus on *making information more transparent and more understandable*, thereby reducing ambiguity and cognitive load for the user [553]. They include translating complex information into simpler forms [553, 393, 410, 253], explicitly mapping options to outline potential outcomes [78, 273], and *using visualization techniques to make choices more salient* [191, 284, 553].

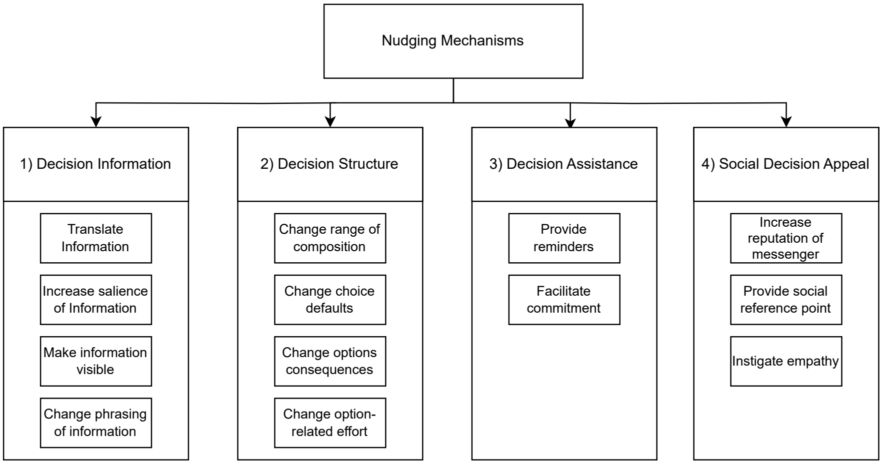


Figure 4.1: Classification of nudging mechanisms as reported in the systematic literature review of Jesse and Jannach [269]. The digital interventions tested in this study use some aspects of decision information (Diversity Information), and social decision appeal (Priming).

4

Additionally, this category involves customizing information to individual user needs [273, 393], making external information visible [100, 212, 556, 410], and providing comparative data to facilitate more informed decisions [553]. Other tactics include using checklists to help track decision progress, providing feedback on user performance, and offering alternative options that may not have been initially considered [273, 328, 347, 393]. Each of these mechanisms subtly influences user choices by altering how information is presented and perceived in digital environments [269].

Decision Structure includes many mechanisms that focus on *arranging options*, impacting how choices are presented and structured. These mechanisms influence how choices are presented to users, subtly guiding their decision-making processes. The arrangement of options can significantly impact user perceptions and choices, often leveraging cognitive biases and heuristics. For example, changing the ease and convenience of options makes some choices more accessible than others, affecting the likelihood of their selection [553]. The order in which options are listed can influence user choices, as items perceived first are intuitively of higher importance [553].

Similarly, changing the physical effort required for choice-making can make them more or less attractive [253, 410]. Splitting options into categories can strategically influence decision-making, such as segregating healthier food options into more diverse categories [273, 410]. Creating friction in decision-making processes can minimize intrusiveness while effectively altering behaviour [78].

By structuring choices in a particular manner, these nudging mechanisms can emphasize or de-emphasize certain options, directing users towards specific decisions without restricting their freedom of choice. This approach is efficient in online platforms where users have many options and require guidance to make optimal decisions.

Decision Assistance comprises several mechanisms divided into two sub-categories (direct and indirect) to support decision-makers in achieving their goals. Decision assistance is crucial in helping users navigate complex decision-making processes, particularly in digital environments where choices can be overwhelming or intricate. Following, we explain how direct and indirect decision assistance mechanisms compare.

1. *Direct mechanisms* directly assist decision-making. This can include tools or features that simplify complex data, provide step-by-step guidance, or offer interactive assistance to help users understand their choices better [78, 253, 333, 563, 599]. For instance, interactive decision trees or guided wizards that lead users through a series of decisions can be part of this sub-category. These tools *reduce cognitive load and make the decision process more manageable and less daunting.*

2. *Indirect mechanisms* indirectly support decision-making by *creating a more conducive environment* for making informed choices. This might include ambient features that reduce distractions, provide a calming interface, or present information in a way that's easier to digest and understand [78]. It could also involve personalized settings that adapt the decision environment to the user's preferences or past behaviour, making the process more intuitive and user-friendly [253, 333, 563, 599].

Social Decision Appeal includes several mechanisms distributed across three sub-categories, focusing on leveraging decision-making's emotional and social aspects. This category subtly taps into users' inherent social nature and emotional responses to guide their choices in digital environments.

1. *Social Proof and Conformity* involves mechanisms that influence peer behaviour and social norms, such as showing the popularity of choices among similar users or highlighting testimonials from respected individuals. This nudges users towards choices perceived as socially acceptable or popular, leveraging the natural human tendency to conform to group norms.

   For example, the *Argumentum-Ad Populum* mechanism (i.e., erroneous reasoning that asserts something is true or beneficial simply because it is a widely held belief) [191] and following the *herd norms* (i.e., individuals in a group acting similarly, collectively, without a central figure guiding or directing their actions) [333, 396, 553] are instances of this.

2. *Emotional Engagement and Resonance* involves mechanisms that connect with users emotionally, such as emotive language, storytelling, or visual imagery that elicits specific emotional responses like empathy, joy, or concern. These nudges can make choices more appealing or relatable to users by appealing to emotions. For instance, instigating empathy with characters [78] and invoking feelings of reciprocity [78] are methods to achieve this.

3. *Community Building and Collaboration* involves mechanisms designed to foster community, cooperation, or user competition. For example, features that encourage group discussions, collaborative decision-making, or gamified elements

that introduce a competitive aspect to decision-making [106]. These nudges can influence users to make decisions that align with a community's or group's goals or values.

These mechanisms demonstrate the diverse ways digital nudging can harness social dynamics and emotional connections to influence decision-making in digital platforms. In the next section of the theoretical framework, we look into ways to leverage the effectiveness and power of digital nudging interventions for social good, such as promoting diversity and inclusion in online workspaces.

4.3.3. Nudging for diversity and inclusion

Digital nudging is present in numerous contexts, such as e-commerce, sustainability and well-being. However, in the context of inclusion and diversity in the online workplace, it is still a domain largely unexplored with a limited – yet growing – number of user studies. Following, we provide some research domains in which digital nudging has been researched to impact diversity and inclusion in online digital spaces.

Nudging for more inclusive sharing economies. The study by Pahuja and Tan [437] proposes a digital nudging approach to reduce racial discrimination in sharing economy platforms like Airbnb. Their work focuses on designing guest profiles that emphasize attributes other than ethnicity, such as hobbies and education, to shift attention away from automatic racial stereotypes. Their preliminary results showed that the most effective nudging intervention was the one that made non-ethnic attributes more salient in the guest profiles. This approach aimed to shift the focus of hosts from automatic racial stereotypes to other shared characteristics between the host and the guest. The study found that this digital nudge could lead hosts (e.g., crowdsourced innovation project requesters) to categorize their guests (e.g., crowd workers) based on shared interests and qualities rather than their ethnic group.

Nudging for more inclusive and equal online mass deliberation. In their study on the design of systems for online mass deliberation and its implications for inclusion, equality, and bias, Shortall et al. [525] identified several themes. These are 1. argumentation tools, 2. automated and human facilitation, 3. gamification, 4. anonymity versus identity, 5. synchronous versus asynchronous communication, and 6. information presentation. Despite its downsides (e.g., superficial engagement, emphasis on competition, manipulation and gaming the system), gamification emerged as a promising tool to increase participation and engagement. Techniques such as rewards, challenges and missions, turn-taking, and feedback and recognition were among the most beneficial interventions. Shortall et al. [525] also highlighted the trade-offs between anonymity and identity in discussions. While anonymity can create a more egalitarian environment and encourage honest expression, it may reduce accountability and civility. Conversely, reducing anonymity increases transparency but can negatively affect engagement and raise privacy concerns. Their work emphasizes the influence of information on deliberations. It shows the need for inclusive and diverse design methodologies since biases in design can emerge from various sources, including cultural assumptions, technical constraints, and the values of the platform developers. Finally, the findings call for an interdisciplinary approach combining computer science, social sciences, psychology, and design to create more effective, inclusive deliberation platforms.

Nudging for more inclusive citizen engagement. The study from Van den Berg et al. [579] conducted in the Netherlands focused on online participation platforms used by the government for citizen engagement. Their research focused on analysing the impact of recruitment messages on participation in these platforms, particularly on gender and age. The study revealed no significant difference in participation rates between women and men, indicating that online platforms can be gender-inclusive. However, age was found to be a significant factor, with participation increasing with age until around retirement age, after which it levelled off. Interestingly, the study found that recruitment messages emphasizing descriptive social norms (highlighting neighbourhood participation) reduced overall participation, especially among senior citizens. This suggests that while recruitment messages can influence participation, their effectiveness varies depending on demographic characteristics. The research highlights the importance of tailored communication strategies to facilitate inclusive participation on online government platforms, emphasizing that different socio-demographic groups may respond differently to various messaging approaches.

Literature-informed nudging interventions for crowd teams diversity. Summarizing the findings from the abovementioned studies on digital nudging interventions for diversity and inclusion, we define some of the rationale behind our proposed study design.

- *Choice of profiling attributes.* The work of Pahuja and Tan [437] demonstrates the effectiveness of emphasizing non-ethnic attributes to shift focus from stereotypes, highlighting the importance of personalization in digital nudging. In our work, we work with ethnic-related attributes (as they are still prevalent in crowdsourcing innovation projects platforms) but combine them with other non-ethnic user profiling characteristics such as work experience and educational level. Displaying ethnicity in user profiles can inadvertently lead to less diversity in choice-making. Furthermore, other studies (e.g., [329, 115]) suggest that the visibility of ethnic and cultural backgrounds can trigger a preference for homogeneity, as individuals often gravitate towards those with similar backgrounds, potentially undermining efforts to promote diversity in team formation and decision-making processes. While inferring the ethnic background of the user may sometimes be difficult on person-to-person recommender systems (e.g., due to the presence of profile photos [265]), focusing deliberately on non-ethnic attributes may be one of the most effective ways to prevent ethnic- and racially-motivated anti-social and exclusive behaviour. Given these findings, our teammates' profiles did not show the ethnic attributes – however, they did show the profile photo. This way, the digital environment used in the study design was intended to represent a familiar team formation system (with profile cards, photos of the teammates, and their attributes) while not explicitly labelling the ethnic identity of the teammates.
- *Choice of Identity in Profiling Crowd Workers.* The research by Shortall et al. [525] highlights the nuanced balance between anonymity and identity on digital plat-

forms, especially in how information presentation impacts user interactions. This insight is particularly relevant to our study's approach to profiling crowd workers. Instead of opting for anonymity, we deliberately design identifiable profiles, allowing each user's unique characteristics to be visible. This decision is rooted in our aim to assess how specific nudging interventions influence decisions when interacting with clearly defined individual profiles. Our strategy aligns with the growing necessity of open innovation projects for traceability and accountability. By ensuring that collaborators are identifiable, our approach closely resembles other online crowdsourcing platforms for open innovation [316].

- *Choice of personalization of nudging interventions.* The work of Van den Berg et al. [579] reveals the nuanced responses of different demographic groups to digital messaging, emphasizing the importance of tailored communication strategies. Our study strongly considers personalization when designing and evaluating digital nudging interventions through Diversity Information and Priming. Finally, these collective insights inform our study design, guiding us to prioritize *socio-demographic personalization and visualization* in our digital nudging interventions to promote diversity and inclusion in crowd-sourced environments.

4.3.4. Priming - implicitly nudging

Priming is the use of initial stimuli to condition individuals before a task. Controlled stimuli are designed to *prime* one's behaviour to act in a certain way. Priming can be either subliminal (the subject is unaware of being primed) or informed (acknowledged by the subject).

Subliminal Priming. Subliminal Priming occurs when the Priming stimulus is presented, so the individual is unaware of it. The stimulus is usually brief and subtle, often below the threshold of conscious perception [483]. For instance, a person might be exposed to a quick, almost imperceptible image or word. Despite being unaware of this exposure, the individual's subsequent behaviour or choices can be influenced by this stimulus [370]. Subliminal Priming works subconsciously, subtly shaping thoughts, feelings, or behaviours without the individual's explicit awareness of the source of this influence [529].

Informed Priming. Informed Priming occurs when the individual is aware of the Priming stimulus. They know they are exposed to specific information or cues influencing their behaviour or thought processes [174]. For example, before performing a task, a person might be deliberately shown images or words related to the task, with the understanding that this exposure is meant to influence their performance. In this case, the individual is conscious of the Priming and can understand its intent, although the effectiveness of the Priming can still vary [38].

Priming for diversity. Our research concentrates on Informed Priming, specifically employing *conceptual stimuli that contextualise information to elicit positive associations with diverse individuals.* This approach guides perceptions and encourages inclusivity through consciously presented content. By exposing crowd users to diversity as explicit Information and positive representations of work culture, we expect them to favour diversity while searching for teammates online. Priming for diversity takes different shapes: from drawing attention to historical injustices to making people recognize their implicit biases while motivating them to act more ethically [580, 122]. While in the context of diversity, implementing Priming techniques has only been hinted at in the past [254, 202], we intend to enlarge the discussion by looking specifically at two main Priming techniques designed to increase diversity: *exposure to counter-stereotypes* and the *All-Inclusive Multiculturalism approach.*

The choice of these two techniques is further justified by the need to expand the discussion on diversity in online team assembly. While previous research has hinted at the potential of Priming for diversity [254, 202], there is a gap in exploring specific, actionable Priming techniques designed to increase diversity. By focusing on exposure to *counter-stereotypes* and the *All-Inclusive Multiculturalism* approach, this research aims to provide much-needed grounding and justification in applying Priming techniques to foster diversity in online environments. This aligns with several other research (e.g., [570, 601, 321, 60]) findings on the role of Priming in motivating ethical behaviour and recognizing implicit biases, further supporting our choice of these techniques for a more nuanced and realistic approach to enhancing diversity in crowdsourcing team assembly.

Counter-stereotypes expose people to positive examples from minority groups [467]. This technique is known to be particularly effective at curbing biases and stereotypes [108, 51, 123, 124, 177, 186, 213]. An example of an effective counter-stereotype is displaying images of female scientists in STEM textbooks. Female students showed higher comprehension of science lessons after exposure to this counter-stereotype compared to reading texts with gender-stereotypical images of male scientists [205]. Even exposing people to the thought of counter-stereotypes has been seen to compel them to abandon the use of categorical labels [230, 255, 312], and develop cognitive flexibility and creativity [312, 198]. Counter-stereotypes benefit not only the reduction of one's access to stereotypical thoughts [230] but also to reduce stereotype threats, meaning the pernicious effects that stereotypes have on the performance of the subject of stereotypes [116]. Exposing individuals to counter-stereotypical images of underrepresented groups can trigger positive automatic associations that can increase positive feelings toward diverse cultures, ethnicities, and genders [14].

All-Inclusive Multiculturalism (AIM) mitigates the effects of stereotypes by explicitly mentioning both majority and minority groups [458, 266]. AIM is also a response to the limitations faced by the two most common initiatives against stereotyping at work: ethnic colour blindness and multiculturalism. While ethnic colour-blindness often unintentionally perpetuates stereotypes by ignoring cultural differences, multiculturalism, though well-intentioned in celebrating diversity, can sometimes inadvertently create a sense of otherness or 'tokenism' among minority groups. This phenomenon occurs when the focus on cultural differences unintentionally reinforces separateness rather than inclusion. To circumvent the feelings of exclusion that other organizational members might feel in the face of multicultural and colour-blind agendas, AIM proposes that diversity should include *all* employees [540]. An example of AIM is explicitly affirming the inclusion of non-minorities (e.g., Dutch workers in a Dutch company) within a general multiculturalism ideology [266]. On the one hand, AIM celebrates differences between individuals and social groups, acknowledging minorities; on the other hand, by explicitly mentioning the essential role that non-minorities play in the workplace, it limits feelings of exclusion and preferential treatment [540].

Combining Counter-stereotypes and All-inclusive Multiculturalism offers a robust Priming strategy for enhancing diversity in online team assembly. Exposure to counter-stereotypes actively challenges and reshapes existing biases by promoting positive attitudes towards underrepresented groups. Simultaneously, AIM complements this by including majority demographics, thereby preventing feelings of exclusion often associated with traditional diversity efforts. This synergistic approach addresses the underrepresentation of diversity and fosters a more inclusive and holistic narrative. Such a deliberate choice in our Priming mechanisms is grounded in balancing broad representation with inclusivity, ensuring all demographic groups are acknowledged and valued. As evidenced in prior research, this strategy is anticipated to encourage more ethical behaviour among crowd users, heightening awareness of implicit biases and inspiring choices that reflect a comprehensive understanding of diversity.

4.3.5. Diversity Information - explicitly nudging

While users cannot be forced to choose diversely, and companies and systems' owners should refrain from censorship, the choice of how and what information gets displayed through interfaces can greatly affect decision-making processes [234]. In addressing the issue of diversity in crowdsourcing team assembly, we explore the potential of nudging through explicit information. This strategy involves disclosing or highlighting relevant information to modify users' awareness and guide their decision-making towards a specific outcome [269]. This approach aligns with findings that suggest well-presented information can significantly influence user choices and behaviours [269]. For this study, we focus on two specific techniques of explicit information: *exposure of attributes* and *recommendations*. The former involves displaying user information to highlight diversity aspects, which affects decision-making by making specific attributes more salient [269]. The latter recommendations, borrowed from person-to-person recommender systems, have demonstrated effectiveness in nudging consumer behaviour in various online contexts [269]. Recommender systems inherently employ several nudging mechanisms, such as information customisation and simplifying choices pertinent to our study's objectives (Section 4.3.2). These techniques are further justified by the mixed results observed in previous studies utilizing different nudging mechanisms, suggesting a need for focused research in this area [269]. By examining *exposure of attributes* and *recommendations* within the specific context of crowdsourcing team assembly, this study aims to contribute to understanding how targeted information presentation can effectively foster diversity. Existing literature indicates significant potential for impact [269].

Exposure of attributes presents and frames information addressing online social stereotypes and users' homophilic tendencies [491, 534]. Gómez-Zará et al. [202]'s work demonstrates that Diversity Scores within teammates' recommender systems can disadvantage diversity, as collaborators favour others similar to them, more so than in scenarios where no Diversity Scores are given. Aside from these results [202], very few other researchers focus on diversity when looking at the repercussions of the exposure of users' attributes on their diversity choices.

However, displaying attributes highlighting racial minority and cultural facets might not always be detrimental to diversity choices[2]. Walker et al.'s study on online hiring decisions shows that personal references such as video testimonials were fruitful at yielding more diverse employees [591].

Regarding racial cues, Walker et al. demonstrate that recruitment websites containing racial diversity cues were more extensively browsed and remembered – particularly by Black[3] participants – than those that lacked diversity references [592]. In contrast to previous studies, our research specifically examines the impact of displaying Diversity Information in the unique setting of open collaboration through crowdsourced innovation projects. While the study by Gómez-Zará et al. provided valuable insights into the effects of displaying DI in a general context, our study dives deeper into how these indicators influence decision-making in environments characterized by open and collaborative innovation. Crowdsourced innovation projects inherently promote a more dynamic and interactive form of collaboration, where participants do not just choose team members but also engage in continuous, open-ended innovation processes. This context may amplify the effects of Diversity Information, as participants might perceive a greater need for diverse perspectives to fuel innovation. However, it could also lead to different challenges compared to more static team assembly scenarios studied previously. In light of this, we posit that the impact of displaying DI in crowdsourced innovation projects might differ from the findings of Gómez-Zará et al.. Their study observed the effects in a more controlled and less interactive setting, but our focus on crowdsourced innovation projects introduces the variable of open collaboration, which could mitigate or exacerbate the adverse effects of displaying DI.

Recommendations are personalized suggestions given to users, which, in our case, are teammates with diverse attributes compared to the user. Recommendations are common among online dating websites (and mobile apps), where users get recommended to matches based on their preferences (content-based filtering), their similarity of choices (collaborative filtering), their similarity of attributes (demographic filtering), or a combination of those. Recommendations' prominence (how bigger or brighter they are compared to the rest of the options) and their position in the list (ranking) are two of the most common methods used when presenting highly recommended choices [301]. Using explanations is also extensively used in person-to-product recommender systems [567], albeit less so for person-to-person recommendations [300]. The use of explanations in reciprocal environments, such as recruitment sites or dating apps, has

[2] In their study, Walker et al. manipulated four employees on a hypothetical organization's recruitment website as either all White (no racial diversity cue) or two White and two Black (racial diversity cue) whilst holding the gender ratio constant (2 men and 2 women).

[3] The term Black refers to individuals of African descent, capitalized to recognize the distinct cultural, historical, and socio-political experiences of Black communities [89, 169]. Furthermore, we used ethnic labels to classify the dummy profiles and the participants. However, these labels were not explicitly printed on the teammates' profiles. Although we needed to rely on over-simplified categories of attributes from the literature (e.g., [422]), we are entirely aware of this approach's significant limitations.

Table 4.1: Mapping of the main Research Question (RQ2) to the specific Research Questions (RQ2.1, RQ2.2, RQ2.3), Hypotheses (H1, H2, H3, H4), and Conditions (Priming, Diversity Information, Priming + Diversity Information).

Main Question	RQs	Hypothesis	Condition
RQ2	**RQ2.1**	*H1: Priming leads to more diverse teams*	Priming
	RQ2.2	*H2: Diversity Information leads to less diverse teams,*	Diversity Information (DI)
	RQ2.3	*H4: Priming and Diversity Information lead to more diverse teams*	Priming + DI

been noticed to be as persuasive as the order of the presentation of the items [300]; this is particularly true when the costs associated with making that choice is significant to that user (in the form of a monetary or emotional investment).

4.4. Research Questions and Hypotheses

Table 4.1 shows the mapping of the Research Questions, the hypotheses, and the conditions chosen for this study design. The overarching aim of our study is to explore the impact of digital nudging techniques on promoting diversity in self-assembled crowd project teams. We have refined our primary Research Question and formulated several specific hypotheses to provide a focused examination of this broad area. These research approaches and hypotheses focus on two digital nudging techniques: Priming and Diversity Information (DI). We believe these techniques are especially pertinent due to their direct applicability in influencing team assembly in crowd-sourced environments, offering a nuanced approach to affecting user choices and thus fostering diversity and inclusivity in team composition. Our primary Research Question is:

RQ2: How do Priming and Diversity Information affect the diversity of crowd users' team member choices?

This question guides the overall focus of our study, directing attention to the potential impact of digital nudging techniques on the diversity of team composition in crowd-sourced projects. To further examine this main question, we have subdivided it into sub-questions, each accompanied by a corresponding hypothesis:

1. **The Impact of Priming.** With this sub-research Question, we aim to focus on the effects of Priming. Priming has been extensively studied in psychology for its efficacy in influencing subconscious decisions and behaviours. However, its application in nudging online workers towards making more diverse choices in

collaborator selection remains under-explored. We propose the following sub-research Question and relevant hypothesis to address this gap:

- **RQ2.1: (How) does Priming affect the diversity of the members that crowd users select for their team?**
- *H1: Priming crowd participants with counter-stereotypes and All-Inclusive Multiculturalism leads them to select more diverse team members.*

  This hypothesis is based on recent academic findings showing that All-Inclusive Multiculturalism can increase perceived inclusion and support for diversity efforts [266]. Combined with team management theory, this approach can help dissolve barriers and create unity in a multicultural workforce [516]. Furthermore, leaders' benevolent paternalism (a core value of nudging and choice architecture) can mitigate the negative impact of intercultural diversity on communication openness in diverse teams [350]. Therefore, Priming crowd participants with counter-stereotypes and an all-inclusive multicultural approach could indeed lead to the selection of more diverse team members.

Through RQ2.1 and H1, we aim to test the effectiveness of Priming, specifically through counter-stereotypes and All-Inclusive Multiculturalism, in enhancing the diversity of team member selection in crowd-sourced projects.

2. **The Role of Diversity Information.** This sub-research Question delves into the influence of explicitly presented Diversity Information (DI) on team selection processes. While there is a growing body of research on the impact of information display on decision-making, the specific role of DI in shaping team diversity in online environments is not well understood. Hence, we investigate the following:

   - **RQ2.2: (How) does displaying Diversity Information (DI) affect the diversity of the members that crowd users select for their teams?**
   - *H2: Displaying explicit Diversity Information (DI) leads crowd users to select less diverse team members.*

     The second hypothesis is based on findings highlighting the paradox of diversity-related information. While intended to enhance team diversity, its display can inadvertently result in the selection of less diverse teams [202]. This effect is particularly pronounced with information on national diversity, which may reinforce social categorization, ultimately impeding inclusive behaviour [120].

   Through RQ2.2 and H2, we intend to explore whether the overt display of DI influences team selection, potentially leading to less diverse outcomes.

3. **Combined Effects of Priming and Diversity Information.** The final aspect of our research seeks to understand the synergistic effects of combining Priming and DI techniques. While both strategies have been individually explored, their combined impact on enhancing diversity in team member selection is unclear.

This integrative approach could offer new insights into effective digital nudging strategies. We therefore propose the following question and hypothesis:

- **RQ2.3: (How) does the combination of Priming and Diversity Information (Priming + DI) affect the diversity of team members that crowd users select for their teams?**
- *H4: Priming crowd participants with Diversity Information (Priming + DI) leads users to select more diverse team members than no Diversity Information and Priming.*

  Our final hypothesis suggests that digital interventions, specifically through Priming and the provision of Diversity Information, will enhance team diversity more effectively than scenarios lacking these interventions. We anticipate that the beneficial impact of diversity-focused Priming will surpass any potential drawbacks associated with presenting Diversity Information. Essentially, we predict a net positive effect on team diversity due to these digital strategies.

Through RQ2.3 and H4, we propose integrating our choice of Priming techniques with Diversity Information. We seek to understand whether combining these two manipulations leads to more effective nudges in promoting diversity in team member selection than when these techniques are used separately.

4.5. Study Design

The task. The task asked participants to form teams for a collaborative effort to create a new coffee slogan for Xpresso (see Table 8.4 in the Appendix). The instructions were as follows: '*We are Xpresso, a coffee company looking for a new company slogan. We need fresh ideas, so we decided to outsource this project. Your task is to select two team members from a list of previously registered individuals to form a team with whom you will collaborate on this project. [...].*'

This task required participants to select team members and envisage working collaboratively with them on the slogan creation. The selection process was critical, as it determined the composition of their collaborative team. This approach, inspired by previous research on crowdsourced team formation [355], aimed to study the impact of diversity indicators and Priming on team selection in a collaborative context.

4.5.1. Participants

The study drew from a sample of 150 people, of which 30 were excluded. The excluded participants belonged to the control condition, five to Diversity Information (DI), seven to Priming, and eight to Priming+DI. Criteria for exclusion were: 1) incomplete submission, 2) incorrect answers to all manipulation checks, and 3) evident lack of engagement (only clicking on the top of the list). Although costly, the latter criteria were intended to exclude outliers from the results from those participants who did not browse through the whole list of teammates.

4

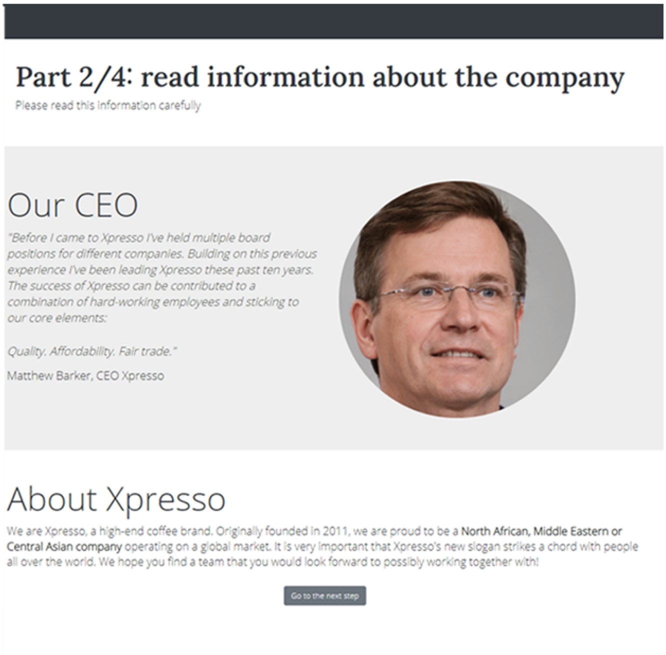


Figure 4.2: Control and DI (Diversity Information) conditions. The page features a standard information page with a white male CEO, representing a typical corporate scenario without explicit diversity cues.

With the intent to capture a diverse pool of crowd workers, we hired participants from two of the most popular online crowdsourcing platforms, namely Amazon Mechanical Turk [441] (*n*=57) and Prolific [447] (*n*=60). The remaining 3 participants were recruited via personal invite. Most participants were from Western Europe (*n*=38), North America (*n*=29), or South Asia (*n*=27). Others were from Eastern Europe (*n*=9), Southern-Europe (*n*=6), South-East Europe (*n*=4), South America (*n*=2).

Only one participant was from the remaining zones of origin [4]. The sample was predominately male (*n*=77). All participants provided informed consent and received 5 USD[5].

4.5.2. Research design

Figure 4.4 illustrates the study design procedure, which we describe later. Ultimately, participants were asked whether they perceived the task and the company as fictitious (as a form of control). Finally, in the end, they were informed that the experimental scenario was fictional and no further steps were needed. For the between-subjects

[4] The zones of origin are intended as geographic regions and do not represent the participants' race.
[5] The average payment for crowd-sourced work [98] and meeting ethical minimum wage requirements.

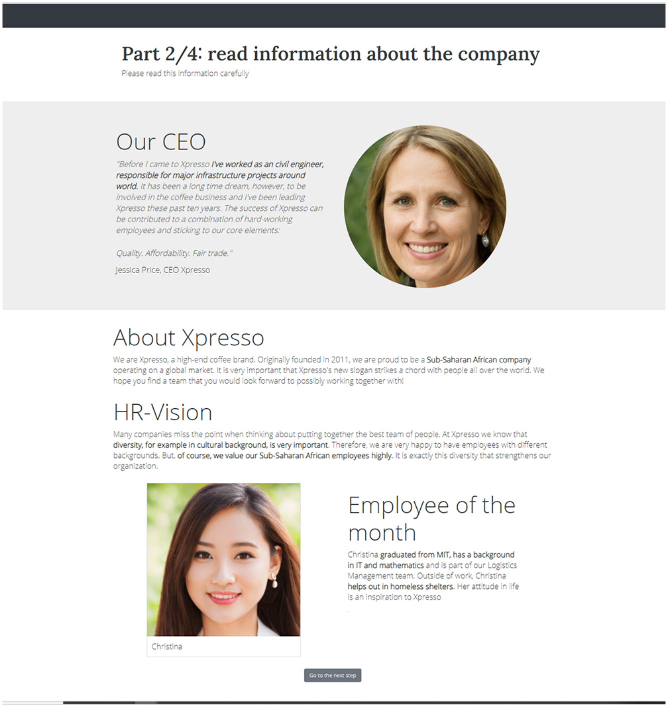


Figure 4.3: Priming and Priming + DI conditions. The information page has been adjusted to reflect the demographics of crowd users, introducing elements of diversity and inclusivity.

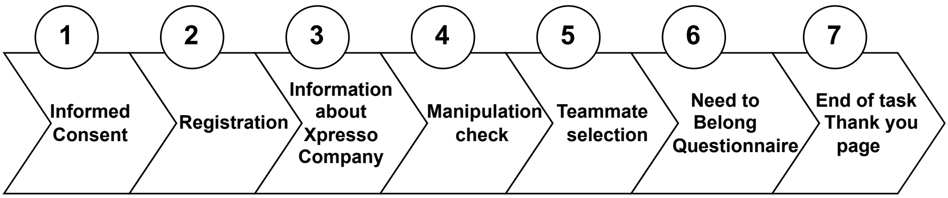


Figure 4.4: Procedure steps of the study design comprising of seven steps.

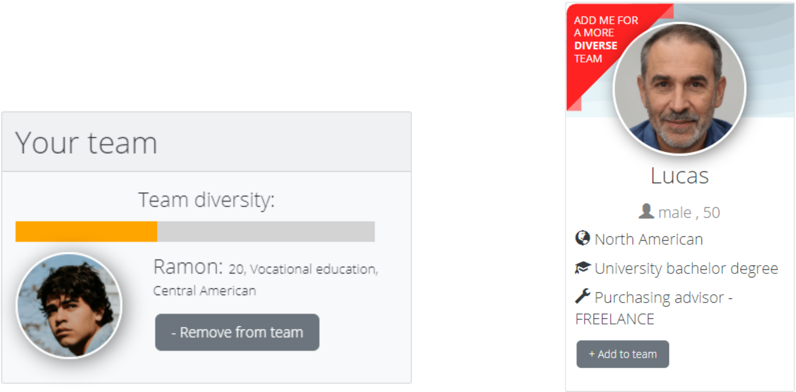


Figure 4.5: Overview of the teammates' profiles showing the Diversity Score as a progress bar (left fig.) and explicit recommendation of a teammate in the form of a red banner (right fig.) on the top-left side of the profile. Two diversity nudging interventions are part of DI and Priming + DI.

study design, we used a 2x2 factorial approach. Participants were randomly assigned to one of the conditions. The factorial design allowed us to observe the independent and interaction effects of Priming and displaying DI on teammates' choices [91]. The independent variables were **Priming** and **DI**; each could be present (Applied) or not (None). The factorial design resulted in the following conditions.

Control condition

The control condition (Condition 1) refrains from displaying a Diversity Score or implementing diversity-related Priming techniques. Instead, it provides participants with some general information about a male, white CEO, identified as Matthew Barker (as illustrated in Figure 4.2 and Figure 4.3), and offers insights into the Xpresso company. Notably, this condition's information page contains dynamically generated content. One significant aspect of this dynamic content is that the region in which the company is located aligns with the region of origin of each participant. This means that every participant, regardless of their region of origin, perceives themselves as part of the majority demographic within the context of Xpresso. This strategic alignment of regions of origin with the company's location serves a crucial purpose—it validates the in *ceteris paribus* assumption[6]. By ensuring that all participants consider themselves part of the majority demographic at Xpresso, we create a consistent baseline for comparison across conditions. This consistency is vital for accurately assessing the impact of Diversity Information and Priming techniques on participants' decision-making processes and behaviours.

[6] *Ceteris paribus* is a Latin phrase that means *all other things being equal* or *holding all other factors constant.* It is used in research and analysis to isolate the impact of a specific variable or condition while assuming that all other relevant factors remain unchanged [506].

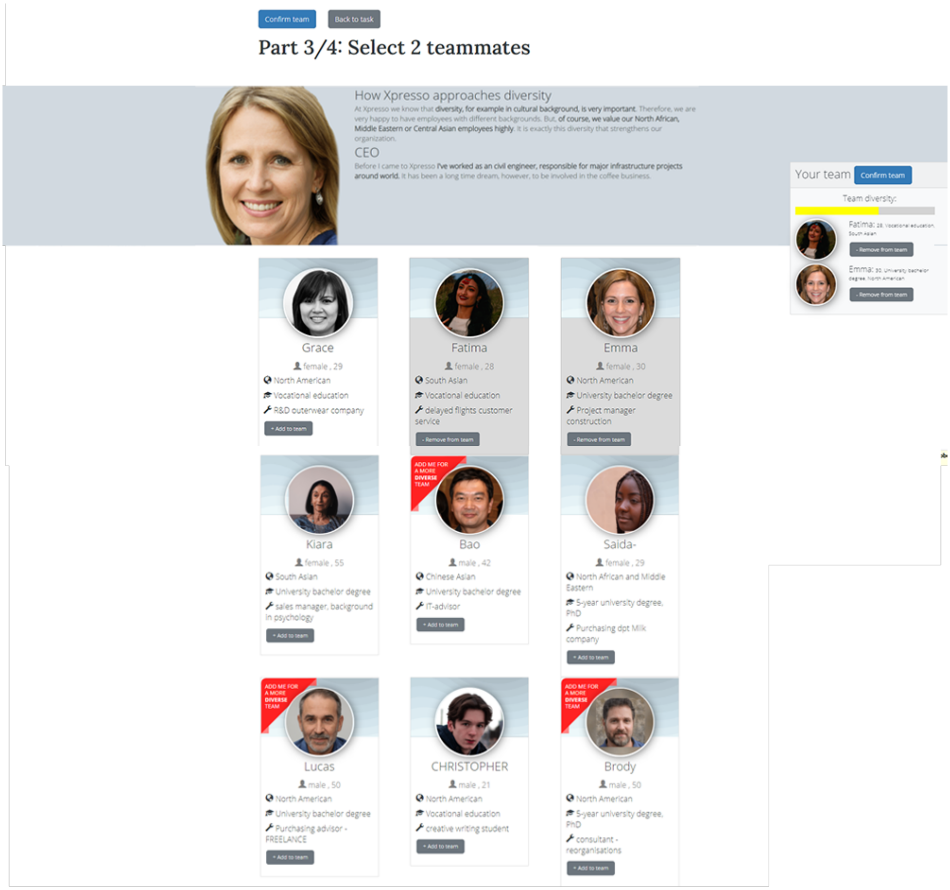


Figure 4.6: Overview of the combination of Priming and Diversity Information as seen by the participants. The interface shows how Priming (displayed at the top of the page) is combined with explicit recommendations (red banners on teammates' profiles) and Diversity Information (the diversity bar shown on the right-hand side). Participants are exposed to conceptual stimuli such as alternative role models and cultural inclusiveness, which are intended to subtly influence their subsequent decisions, encouraging the choice of team members with diverse attributes.

4

Diversity Information Condition

In the Diversity Information condition, while maintaining the basic structure of the control group, we introduced two critical modifications on the team selection page to emphasize diversity. These modifications were informed by existing literature highlighting the effectiveness of visual cues and real-time feedback in influencing user decisions (cite relevant studies).

Progress Bar for Diversity Display. Upon adding a member to their team, participants encountered a progress bar visually representing the team's diversity level. This bar displayed the team's diversity as an aggregated measure, using the Blau score on a scale from 1 to 100 (illustrated in Figure 4.5a). The Blau score, a widely recognized metric in diversity research [52], calculates the probability that two randomly selected team members will differ regarding specific attributes (e.g., gender, ethnicity, background). By converting this score into a percentage, we aimed to provide an intuitive and immediate understanding of the team's diversity level, encouraging participants to consider diversity actively in their selections—more on how the Blau score was calculated in Section 4.5.4.

Recommendation Banners for Increased Diversity. The second intervention involved banners recommending dummy profiles with the tagline: *Add me for a more diverse team* (see Figure 4.5b). These recommendations were triggered when adding a suggested profile could increase the team's Diversity Score above 75%. The Diversity Information (DI) scores were dynamically adjusted based on the user's attributes and already chosen teammates. This approach is grounded in research suggesting that direct suggestions can effectively nudge users towards more diverse choices [282, 269, 78]. These interventions were designed based on the hypothesis that real-time, visually engaging feedback could influence participants' decision-making towards creating more diverse teams [79]. The Blau score was selected as the diversity measure due to its robustness and widespread acceptance in social science research [52], providing a quantifiable and meaningful way to assess diversity. Using these mechanisms, we investigated whether direct, informative cues could alter participants' natural inclinations and lead to more diverse team compositions, in line with the objectives of promoting diversity and inclusion in team-building contexts.

Priming Condition

Within this condition, we use the concept of Priming as a pivotal experimental technique. In psychological research, Priming involves subtly exposing individuals to specific stimuli (e.g., alternative role models, cultural inclusiveness) that can unconsciously influence their responses to later stimuli [294]. In our study, this entails presenting participants with conceptual stimuli that subtly influence their decision-making process, especially in selecting team members with diverse attributes. The Priming condition mirrors the control group in its basic structure. However, it incorporates counter-stereotypical elements (like a female CEO and a minority employee of the month[7]) and All-Inclusive

[7] The minority employee is portrayed as a counter-stereotypical woman in science, with an ethnicity differing from the participant's, to show her minority status in combination with the HR statement

Multiculturalism (AIM) elements (such as an HR vision statement[8]) (illustrated in Figure 4.3). As outlined earlier, the techniques implemented in this experiment phase are forms of Priming. By exposing participants to conceptual stimuli like alternative role models and representations of cultural inclusiveness, we aim to subtly guide their subsequent decisions towards selecting more diverse teams. The images of the CEO and the employee of the month were displayed on the information page, the manipulation checks page, and the team selection page. The HR-vision statement was placed on the information page and prominently at the top of the manipulation check and team selection pages. These Priming interventions were customized to the participants' attributes. For instance, if a participant identified as White, the counter-stereotypical imagery would feature individuals of non-White ethnic backgrounds to enhance the impact of the diversity message.

Priming + Diversity Information (DI) condition

Figure 4.6 illustrates the integration of Priming and Diversity Information (DI) on the teammate selection page. In this condition, we combine the Diversity Information condition (DI) manipulation with the Priming condition by incorporating the counter-stereotypes and AIM characteristics. This means that users in this condition will be exposed to diversity-related information similar to DI while benefiting from the positive associations and inclusivity aspects introduced by the counter-stereotypes and AIM characteristics used in the Priming technique.

4.5.3. Materials

We created 30 realistic-looking teammate dummy profiles (see example in Figure 4.5b). Each of these was assigned the relevant attribute characteristics The dummy profile characteristics were distributed as follows. Age (Generation Z (*n=9*), Millennials (*n=10*), Generation X (*n=11*)), gender (male (*n=14*), female (*n=15*), other (*n=1*)), functional background such as skills, expertise, and professional experience (10 types (*n=3*)), region (Europe (*n=7*), North Africa, Middle East or Central Asia (*n=1*), Latin America (*n=3*), East Asia (*n=3*), South and South-East Asia (*n=7*), Caribeans (*n=1*), Sub-Saharan Africa (*n=1*), North America and Australasia (*n=7*)), and ethnicity (White (*n=12*), Black (*n=4*), Asian (*n=11*), Latino (*n=3*)).

For an overview of the attributes, see Section 4.5.4 and Table 8.7 in the Appendix). The profile pictures were partly AI-generated [283] and partly acquired as royalty-free pictures [576]. Between 30-40% of the photos were distorted or colourized to resemble as closely as possible the level of variance and individuality of profiles that one would expect from real-life matchmaking platforms. The dummy names were common for the region they supposedly came from [319]. Dummy attributes, such as age, region, and ethnicity, were assigned to the 30 profiles based on the population statistics of workers from crowdsourcing platforms like Amazon Mechanical Turk. This approach reflected

[8]HR statement: *"Many companies miss the point when thinking about putting together the best team of people. At Xpresso, we understand that diversity, for instance, in cultural backgrounds, is crucial. Therefore, we are proud to have employees from diverse backgrounds. However, we also highly value our [PARTICIPANT's REGION] employees. This diversity is what strengthens our organization."* This statement was adapted to reflect the participant's region, positioning them within the majority

the predominant demographics of these platforms, notably Indian and North American individuals from the millennial generation [139]. This design choice was intended to provide a *realistic representation of the typical crowd-worker population*, thereby enhancing the ecological validity of the study. By mirroring the actual demographics of platforms like Amazon Mechanical Turk, we aimed to create a more authentic and relatable decision-making environment for the participants. Furthermore, limiting the number of profiles to 30 resulted in a trade-off between representing as many combinations of diversity attributes and ensuring participants could look at and assess all the profiles. This design choice rationale followed the findings from related work on how limiting the number of options could help manage cognitive load, potentially leading to more meaningful and considered decisions by participants [629].

4.5.4. Metrics

Dependent measure

Our metrics are based upon Gómez-Zará et al. [202]'s study design, calculating team diversity as an aggregate measure. Teams consisted of two dummy profiles chosen by participants plus themselves. We chose to study diversity among crowd teams of size three since we wanted to provide an initial analysis of a basic team unit. However, we also avoided studying dyads as often a debated subject in CSCW research on group formation [402].

Team diversity for each attribute was calculated using the Blau index, a widely accepted method in diversity research [288, 202, 52]. The Blau index is formulated as $1-\sum_{i=1}^{k} P_i^2$, where P_i is the proportion of team members in the i-th category out of k, the total number of categories for that attribute [532]. This index calculates a Diversity Score between 0 and 1 for each attribute, serving as a measure for diversity as variety [227]. The final team diversity score is derived from the sum of Blau scores for all included attributes divided by the number of attributes, providing a complete view of team diversity.

Although most diversity traits were categorical, age was categorized into ranges or generations to address the challenge of continuous data creating a sparse matrix [421]. The following example shows how the Blau index calculates the Age Index Score for a team with members from distinct age categories.

Example calculation Age Blau Score. Considering a team of three, each from a different generation, and a total of 5 generations ($K_a = 5$), the process of calculating the Blau Score for the team age diversity is as follows:

1. **Calculate the proportion squared** (P_k^2) for each represented generation:
   - For the three represented generations: $P_k^2 = \left(\frac{1}{3}\right)^2 = \frac{1}{9}$.
   - For the unrepresented generations: $P_k^2 = 0^2 = 0$.
2. **Sum the squared proportions**:

$$\sum_{k=1}^{K_a} P_k^2 = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{3}{9} = \frac{1}{3}.$$

3. **Apply the Blau index formula for age diversity** (D_a):

$$D_a = 1 - \sum_{k=1}^{K_a} P_k^2 = 1 - \frac{1}{3} = \frac{2}{3}.$$

The resulting Blau Score for the team's age diversity is approximately 0.6, a moderate age/generational diversity level for a team of three.

Independent measures

To measure diversity, we used the following independent measures gathered from the participants: **1) Surface-level traits**: age and gender, **2) Deep-level traits**: functional background (a mix of skills, expertise, and professional experience) and level of education, and **3) Surface and deep-level traits**: cultural background. The sub-categories are shown in Table 8.3 of the appendices.

Surface-level (relations-oriented) traits: Age and gender. Age is treated here as a measure of differences of years within a team regarding variety. As a continuous metric, age must first be converted into a categorical variable via discretization [421]. Based on Ferrero-Ferrero et al. [166]'s classification of generations, we categorized the participants and the dummy profiles' ages into five generations: Greatest Generation/Silent (aged 76+ in 2021), Boomers (57-75), Generation X (41-56), Millenials (25-40), and Generation Z (18-24). This grouping helped with the clustering of the subjects in terms of their generational differences, especially regarding their values, trust of authority, and independent thinking [166, 575, 549]. Gender, being by definition categorical, did not require additional discretization.

Deep-level (task-relevant) traits: Functional background and level of education. For the classification of the functional backgrounds, we revised the nine categories by Pegels et al. [448] into the following ten: 1) Information systems, 2) Customer service, 3) Sales and marketing, 4) Engineering, R&D, 5) Purchasing/Procurement, 6) Operations, administrations or manufacturing, 7) Consultancy, 8) HR/personnel, 9) General management, 10) Creative sector. To ensure that the potential combination of functional backgrounds could be observed in practice and be considered sufficiently diverse [421], we revised the list by focusing on what makes each sector distinguishable from one another (i.e., the unique skills and knowledge needed for each). Additionally, manufacturing was added to the revised list as it was not present in the original version by Pegels et al. [448]. To validate the applicability of this classification, we checked these revised categories against Indeed [258]'s list of most popular jobs in the United States as of 2020 [258]. For levels of education, we chose the following: (1) Primary education, 2) High school + Diploma, 3) Vocational education, 4) a University bachelor's degree, and 5) 5-year university degree or PhD.

Surface and deep-level (relations-oriented) traits: cultural background. To capture both surface and deep-level attributes of cultural background, we calculated this metric as the mean cultural background diversity of two other features: ethnicity and region. We classified participants and dummy profiles into one of five categories: Asian, Black,

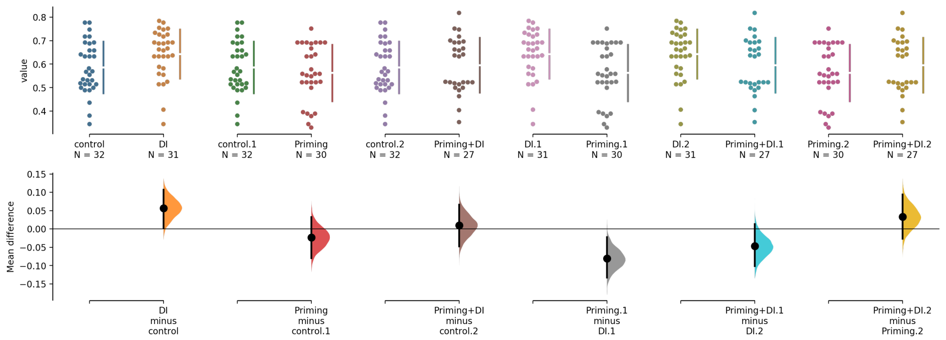


Figure 4.7: The mean difference for 6 comparisons is shown in the Cumming estimation plot. The raw data is plotted on the upper axes; each mean difference is plotted on the lower axes as a bootstrap sampling distribution. Mean differences are depicted as dots; the ends of the vertical error bars indicate 95% confidence intervals. The Diversity Information (DI) produces significantly higher diversity (mean difference) than the other conditions. DI.1 and DI.2 are automatically generated labels during the pair-wise comparison, and both represent the Diversity Information (DI) condition.

Brown, Latino, and White [173, 349][9]. For the calculation of the region of origin, we settled for regional data [103, 513] from the *European Standard Classification of Cultural and Ethnic Groups* adopted by Schneider and Heath [513]. This categorization does not centre around nationalities, as it combines regional and cultural aspects – such as religions – to provide a rather exclusive and complete list that captures more than just surface-level aspects of cultural background [536, 228].

4.5.5. Procedure.

Figure 4.4 shows the procedure consisted of seven steps.

1. *Informed Consent and Task Description.* Participants received detailed information about the study and were requested to give informed consent (see Table 8.2 in the Appendix). Additionally, they were introduced to the task at hand, which involved generating a new slogan for Xpresso, a coffee company, as part of its worldwide advertising campaign (see full description of the informed consent and task in Table 8.4 of the Appendix).

2. *Registration.* Participants then registered (see form entries in Table 8.5 of the Appendix), providing their details (same list of socio-demographic attributes as the dummy profiles' in Section 4.5.3) to commence their study participation formally.

3. *Information about the Requester (Xpresso Company).* In this phase, participants learned about Xpresso, the hypothetical company pivotal to the study. This

[9]This categorization is intended as a starting point for exploring how diverse cultural backgrounds influence team interactions and outcomes. It is a preliminary simplification meant to operationalize the concept of cultural diversity in a research context. Nonetheless, we are acutely aware of the need to incorporate – in future research– more nuanced and self-identified measures of cultural, racial, and ethnic identity [349].

background information aimed to give context to the task and included Xpresso's company profile, its founding year, a message from the CEO, and highlights such as the employee of the month.

4. *Manipulation Check.* Table 8.6 of the Appendix provides an example of the manipulation check used to test participants' attention towards the task and specifically about the outsourcing company. The manipulation questions varied slightly between conditions but always tested for certain essential aspects of the hiring company. Through the questions, we checked whether participants remembered the company's founding year and the CEO's name.

   For those in the Priming experimental groups (Priming and Priming + Diversity Information conditions), we also asked about certain employees mentioned in the company's HR policy. The question about the company's founding year was twofold: it helped us see if participants were paying attention and distracted them slightly from focusing too much on diversity aspects.

   Erroneous answers to the manipulation checks (i.e., all participant's answers were incorrect) were removed from the final pool of contributions. This activity was part of the exclusion criteria (see Section 4.5.1).

5. *Teammate Selection.* Each participant chose teammates from a set of profiles. Participants were informed that Xpresso might select their chosen team for slogan creation, and they could be invited to collaborate. However, they retained the option to decline this invitation. Participants had to pick two team members from 30 fictitious profiles portrayed as real registrants. These dummy profiles, presented randomly, included a photo, username, region, education level, and a job description indicating various professional backgrounds.

6. *End of Task and Thank You Page.* Participants received thanks for their contribution and were compensated afterwards via the crowdsourcing platform.

4.6. Results

Section 4.6.1 broadly addresses the main Research Question (*How does display Diversity Information or Priming diversity affect users choosing more diverse team members?*); it concerns the testing of all four hypotheses listed in the introduction. Section 4.6.2 presents a posthoc analysis and any secondary effects of participants' and dummy profiles' characteristics. We use Kruskal-Wallis tests to analyze independent factors. We look at crowd users' region of origin (Section 4.6.2).

With an unpaired one-tailed t-test, we investigate dummy profiles' attributes that contributed to their popularity (Section 4.6.2). A two-sample paired Wilcoxon test highlighted differences in gender preferences and the presence of gender-driven homophily (Section 4.6.2). Finally, we used a linear regression model to evaluate possible presentation biases in the study design confirmed by dummy profiles' popularity (Section 4.6.2).

Table 4.2: Pairwise comparisons of unpaired mean differences with 95% Confidence Interval and p-value of the 2-sided non-parametric permutation t-test

Condition 1	*Condition 2*	*Unpaired mean difference*	*95% CI*	*p-value*
Control	DI	0.0569	[0.00221, 0.107]	0.044
Control	Priming	-0.0233	[-0.0794, 0.032]	0.417
Control	Priming + DI	0.0098	[-0.048, 0.0662]	0.741
DI	Priming	-0.0802	[-0.133, -0.0222]	0.0056
DI	Priming + DI	-0.0471	[-0.102, 0.0136]	0.114
Priming	Priming + DI	0.0331	[-0.0263, 0.0944]	0.298

4.6.1. Hypotheses testing

Displaying Diversity Information positively affects diversity

Given the non-normality of the data (Shapiro-Wilk normality test, p = 7.166e-4) and the factorial design of the study, we could not run a two-way ANOVA but used a Mann-Whitney U test instead. DI was the only condition to yield statistically significant results (Mann Whitney U=2205.00, p=0.033).

4

This finding implies that the sole display of Diversity Information is sufficient to impact the diverse choice of teammates. Given that both Priming and Priming + DI did not yield statistically significant results (Mann Whitney U=1542.00, p=0.179), our study cannot confirm the potential positive effects of Priming and Priming + DI on diversity. As drawbacks to traditional statistical hypothesis testing have been noted, we also provide an estimation statistics analysis focusing more on effect sizes. Figure 4.7 shows the data from a multiple-two group analysis with estimation statistics[10] [245]. We compare conditions (control, DI, Priming, Priming + DI) and display the results through a Cumming estimation plot of several sets of two-group data, enabling pair-wise comparison of mean differences (Table 4.2)[11]. This confirms the earlier Mann-Whitney result that DI resulted in more diverse teams. Priming resulted in significantly less diverse teams than DI. It seemed to have had a slight negative impact on diversity, which only became significant when comparing it to the condition that had a positive impact.

Additionally, we ran Kruskal-Wallis tests for each separate diversity attribute. We examined whether the treatment conditions significantly affected team diversity independently. The tests showed no significant differences between the control and treatment conditions for gender diversity (p=0.068), ethnicity diversity (p=0.219), age diversity (p=0.242), education diversity (p=0.546) and functional background (p=0.491), except for the region of origin. We found that Priming to the control negatively affected differences in the region of origin (adjusted p= 0.038). Results on the effects of Priming and DI show that although Priming did not positively affect diversity choices, the display of Diversity Information (DI) positively impacted diversity.

[10] This method uses bootstrapping; resampling the distribution of the difference in means approaches a normal distribution, allowing parametric tests to be used.

[11] 5000 bootstrap samples were taken; the confidence interval is bias-corrected and accelerated. The P value(s) reported is the likelihood(s) of observing the effect size(s) if the null hypothesis of zero difference is true.

Five thousand reshuffles of the control and test labels were performed for each permutation P value.

Table 4.3: Kruskal-Wallis: p= 0.005. Comparison of Country of origin and diversity choice

Comparison	*Z*	*P.unadj*	*P.adj*
Participants from Europe - Participants from North-America	2.778	**0.005****	**0.016***
Participants from Europe - Participants from South and South-East Asia	2.424	**0.015***	**0.046***
Participants from North-America - South and South-East Asia	-0.250	0.802	1.000

This result contradicts previous findings [202]. We also observed that through Priming (counter-stereotypes and AIM), users were less likely to choose teammates from other regions.

4.6.2. Post-hoc analysis

The results do not support the hypotheses. They even show some opposite results (Diversity Information enhance diversity choices). We deemed it insightful to provide an extensive posthoc analysis for mainly two reasons: 1) to validate the data by confirming different expected behaviours and 2) to see what other factors affect the choice of team members.

Regions of origin perceive team diversity differently

As the participant pool had highly different demographic backgrounds, we looked at significant differences in team diversity between participants from various regions of origin. Participants from Europe, South Asia, and North America were the majority populations within the participant pool (95%). Those participants were part of the analysis. Participants from South Asia and North America were normally distributed, but the participants from Europe were not (Shapiro-Wilk test for the participants from Europe: p= 0.017). We, therefore, opted for the non-parametric Kruskal-Wallis test (Table 4.3). We compared three regions, with team diversity as the dependent variable. The test showed that the **participants from Europe significantly chose more diverse team members than those from North America and South Asia**.

Next, we examined the effect of the treatment conditions per region. As the participants from Europe yielded significantly different diversity results, we assessed them separately from participants from North America and South Asia. We conducted a two-way ANOVA test that included only participants from Europe. The assumption of normal residuals (Shapiro-Wilk model residuals: p= 0.418) and homogeneity of the variances (Levene's test: p= 0.074) were met in this case. The two-way ANOVA showed that participants from Europe were positively affected by the display of Diversity Information (p=0.013). **Participants from Europe are therefore more likely to choose diverse teammates, and are also more positively affected by Diversity Information than other participants** [12] Conducting a two-way ANOVA including all non-participants from Europe furthermore showed that Priming may negatively impact the choice of more diverse team members (p=.011).

[12] Diversity is calculated as an aggregate measure of all profiling attributes, not only region.

Table 4.4: Linear regression order of appearance (Shapiro-Wilk: p-value = 0.1482).

Coefficients	*Estimate*	*Std. Error*	*t-value*	*p-value*
(Intercept)	13.73103	1.27922	10.734	**1.97e-11** ***
Ranking	-0.36974	0.07206	-5.131	**1.94e-05** ***

Functional background matters when choosing teammates

While creating a coffee slogan for a company does not necessarily require formal education, we did expect to possibly see a preference for teammates with a sales and marketing background[13]. We compared the means of selected team members with and without a sales background. Due to assumed normality and homogeneity of variances (Shapiro-Wilk: sales p= 0.536, not-sales p= .099; F-test: p=.919), we conducted an unpaired one-tailed t-test which showed that, indeed, **dummy profiles with a sales background were significantly more frequently selected than those without one** (t-test: p= 1.57e-4).

Same gender matters when choosing teammates

One of the stronger homophilic tendencies is that of gender. The expected behaviour is that participants choose significantly more team members of the same gender. Due to the unequal distribution of genders among the dummy profiles, we normalized the scores of each participant's same-gender team members and different-gender team members. We ran a two-sample paired Wilcoxon test (due to non-normality, Shapiro-Wilk: p= 3.263e-11) and found a significant difference between selected same-gender team members and selected different-gender team members (upper-tailed Wilcoxon: p=0.002). We repeated the Wilcoxon test for female participants (n=44) and male participants (n=76). Female participants similarly selected significantly more same-gender team members (upper-tailed Wilcoxon: p=0.011). Male participants followed the same pattern (upper-tailed Wilcoxon: p=0.044). These results indicate that **gender homophily is present, regardless of the condition.**

Order of appearance matters when choosing teammates

We examined whether the order of appearance of the dummy profiles influenced the choice of team members. We expected a linear relationship and conducted a linear regression where the x-axis represented the order of appearance (places 1-30). The y-axis represented the selection frequency of the dummy profiles. We assumed normality (Shapiro-Wilk: p= 0.148). We found that the order of appearance of the dummy profiles showed a strong correlation with the selection of team members (p= 1.94e-5). So, **The order of appearance[14] played a role in the choice of team members from a selection of 30 dummy profiles** (Table 4.4).

[13]The different functional backgrounds were equally distributed among the dummy profiles.

[14]The dummy profiles were in a random order for each participant. No specific dummy profile had an unfair advantage to be selected based on this order in the regression analysis.

4.7. Discussion

This study examined how Priming and displaying personalized Diversity Information affects choosing more diverse team members among crowd users in an open collaboration context. We examined whether Priming and displaying Diversity Information together increased crowd team diversity. We conclude that: a) Diversity Information (DI) alone positively affects the choice of more diverse crowd teams; b) Priming alone negatively impacts crowd team diversity when compared with the display of DI; c) combining Priming and Diversity Information (Priming + DI) does not effectively nudge diverse choices of crowd teammates. DI was expected to decrease team diversity [202], yet results indicate no significant drop as participants selected more diverse teammates (especially in the condition showing only Diversity Information). Future work will be needed to disentangle the effect of each DI intervention (the progress bar and the profile recommendations). Priming, which was expected to increase team diversity, yielded no significant positive impact. Results even indicate the opposite effect may occur, especially regarding the homophilic preference of teammates of the same region. There are several possible causes for our diverging results. Comparing our results concerning Diversity Information with the ones from Gómez-Zará et al. [202], we detect study design differences that may have contributed to differences in outcome. In particular, their sample differed significantly from that used in this study. Participants from Gómez-Zará et al. [202] were fewer (N=46, of which the most significant part was American), all volunteering and non-paid undergraduate students. At the same time, we hired a diverse pool of crowd workers (N=120 from three continents) compensated and motivated through a crowdsourced innovation project competition, more in line with a real-world setting. Finally, we suggest repeating the study with other scenarios, such as political writing and analytical problems, to evaluate the effects of different task types.

4

4.7.1. System design recommendations

The findings from the analysis of the three experimental design conditions enabled us to produce the following system design recommendations.

1. *Design platforms that openly explain diversity instead of subliminally.* Our first recommendation is to *design platforms that openly explain diversity instead of subliminally.* Direct diversity interventions like recommendations of diverse teammates and scores are more effective at nudging toward diversity choices than suggestive and indirect means. Our use of Priming interventions might have gone unnoticed by users as it was seamlessly integrated with the rest of the task description. Using UI elements distinctly and concisely can be more effective at nudging than more covert digital Priming techniques. Combining different kinds of nudging techniques does not seem to yield predictable outcomes. It can even risk confounding information by incorporating conflicting perspectives as diversity nudges carry implicit assumptions.

2. *The importance of testing assumptions.* While counter stereotypes and AIM trigger associations to cultural identity and social belonging, combined with

other nudging techniques, they could activate undesirable reactions to diverse choices. Based on our results, we suggest testing before combining diversity nudging techniques into a single system.

3. *Personalization is the future of inclusion and diversity.* Our third recommendation is to avoid overly generalized inclusive statements and references and focus on *designing a personalized workspace adjusting to characteristics and task objectives.* We noted through our work that personalized recommendations and Diversity Scores were more effective than Priming on targeted traits.

4. *Transparency and Accountability in Nudging Interventions.* Within the scope of this chapter, we assign critical importance to transparency and accountability in digital nudging interventions. This is particularly pertinent in initiatives aimed at promoting diversity. The emergent field of explainable AI (XAI) emphasizes the necessity of demystifying complex algorithmic processes for end-users. Research such as that by Rai [473] aligns with this perspective, indicating that user trust in digital services is significantly bolstered when they understand the rationale behind AI's decisions and suggestions. This aspect of user comprehension is paramount in digital nudging for diversity. This work contributes to a body of research advocating for exploring the benefits of making diversity-nudging interventions transparent and accountable. Ultimately, this research seeks to unveil how digital nudging's transparency and accountability (and the lack thereof) can affect user engagement and decision-making. Mainly, it focuses on how different digital interventions- and their more or less explicit ways of nudging- can hinder or foster diverse and inclusive online environments.

4.7.2. Limitations

This study encompasses a range of limitations, each with its unique implications.

1. **Choice and Number of Participants:** The recruitment platforms' diversity and the limited number of participants (*n=120*), dictated by strict quality-control procedures, represent a constraint. Despite manipulation checks ensuring reliability, it was challenging to fully gauge the crowd users' engagement levels and true intents in the task. There's a risk that the participants made decisions hastily to optimize time rather than based on the provided information. Furthermore, the order in which options were presented influenced selections, indicating a tendency to choose earlier options without thorough comparison.

2. **Design of Task and Profiles:** Although participants believed in the reality of the teammates and the outsourcing company[15], no tangible proof was provided to reinforce this belief. We also did not consider factors such as the perceived attractiveness of profile photos. The limitation of the teammate pool to a small number (*n=30*) possibly did not reflect the diversity found in real-world settings, consequently restricting the range of profile combinations available.

3. **Specific Implementations of Diversity Information and Priming:** The study's findings are limited to the effects of displaying Diversity Information and Priming

[15] Validated by crowd participants' feedback that they thought profiles and tasks were genuine.

in a crowdsourcing environment. The implementations used (e.g., All-Inclusive Multiculturalism, Counter-Stereotypes, Diversity Scores with a bar and colour) do not encompass all possible Priming and DI display methods. Other interventions might yield different outcomes. Moreover, active learning interventions like discussions and problem-solving could offer alternative approaches to Priming, which might be more enduring and less susceptible to arbitrary factors.[403] Active learning could involve participants more directly in diversity enhancement, giving them greater responsibility and engagement than subtler nudging techniques.

In summary, while this study provides valuable insights, its limitations highlight areas for further exploration and refinement in future research, particularly in participant diversity, task design, profile distribution, and methodological approaches to enhancing diversity awareness.

4.8. Conclusion

This chapter examined ways digital nudging may improve diversity in open collaboration crowd teams. It shows that designing diversity-enhancing interfaces, particularly for practical implementations, is more context and user-target-dependent. As a form of nudging, displaying Diversity Information surprisingly enhances diverse choices among remote users hired for a crowdsourced innovation project. On the contrary, Priming strengthens homophilic biases toward users' profiles from the same region. Results from testing diversity Priming techniques (AIM and counter-stereotypes) even hint at possible adverse effects on the diverse choice of crowd users. Overall, we also observe homophilic tendencies toward the same gender among crowd users choosing teammates online. Online team formation systems for crowd collaboration have many opportunities to alter their users' perceptions and decision-making processes. Yet, certain types of nudges may trigger adverse reactions toward diversity. Based on our results, displaying personalized Diversity Information seems most promising. In Chapter 5, we evaluate whether and how profiling attributes (surface- and deep-level) impact online collaboration and team performance. Furthermore, we investigate the actual usefulness of attributes (compared to the perceived usefulness studied in Chapter 3).

5

Attributes and Dynamics in High-Pressure Crowd Teams

5.1. Abstract

Chapter 3 showed that crowd workers *perceive* the deep-level trait Personality as useful for team formation. However, it did not test whether using this attribute may actually improve team performance. 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 chapter addresses the Research Question **RQ3: How do personality and communication patterns affect online ad hoc teams under pressure in emergency response situations?** We explore the dynamics between 120 crowd participants in 60 virtual dyads and their collaboration outcomes during 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 used action and response statements more. The team role was linked to the individual's preference for specific communication patterns and related to their perception of 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.

5.2. Introduction

Situations that require working together, fast, and efficiently under pressure are on the rise, especially in an increasingly fragile global ecosystem [309, 512]. From handling widespread geopolitical conflicts [178] to mitigating environmental disasters [187], several organizations are investing in crowdsourcing intervention to aid large-scale mobilization of resources, including emergency shelters and disaster-event detection [628, 454, 539]. Likewise, virtual teamwork enacted in high-urgency, high-stress tasks is in demand. Grassroots social engagement (i.e., Covid-19 pandemic hackathons [107]), incident response squads [440], community response teams, and on-call software solution teams [21] 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 centres in coping with widespread pandemics [161]. Nevertheless, little is known about the factors that can make or break such teams. In this chapter, we attempt to answer the thesis' first question **RQ3: How do personality and communication patterns affect online ad-hoc teams under pressure in emergency response situations?**. More specifically, we tackle questions such as:

a) *What personality characteristics render high-stake online teams successful?*

b) *Which skills, abilities, or socio-cultural elements must be considered when forming these teams?*

c) *Are there any particular communication patterns that can serve as early signals of effective teamwork under stress?*

Answering these questions is crucial to leveraging 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 [310, 292], 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 [150]. Characteristics of team performance in rapid-response, high-stress contexts are team members' ability to work in a team and personality traits [547, 387]. However, studies on high-stakes teams focus 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 [97]. 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.*

5.2.1. 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" [302] puzzle video game. Participants work in dyads, and their shared mission is to defuse a bomb placed within a maze by combining information unique to each of them. One participant is assigned the role of the "Defuser": they can "walk inside the maze towards the bomb and defuse it but do not know where the maze walls are. The other participant is assigned the role of the "Lead Expert: they have the maze map but cannot walk in it. The Defuser and the Lead Expert must exchange information and actions to defuse the bomb within a limited time. The task has been designed to have the same critical characteristics as emergency response tasks: a high-demanding environment, enforced role division, performance pressure and stress.

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 [46], 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 quickly. The typical environmental constraints of high-demanding tasks (time, urgency, risks) command 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.

Enforced role division

During emergency cases, each team member has a distinct and specific role to play [34], 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 [226]. Although role division is typically fixed for these response units (e.g., medical, logistics, security, public relations, etc.), it must 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 the constituents that fundamentally determine the execution of role-based virtual teamwork emergency response.

Performance pressure and stress

Prior work has shown that users involved in games such as the crowdsourcing task exhibit various forms of stress [495] and heightened emotional states [229]. These teams are more susceptible to allostatic load, i.e., the process of "wear and tear" experienced by team players facing stressful conditions [125]. Regarding the definition of stress, there are two kinds of stressful conditions and stressors [360]. 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 gains [630].

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 [498, 185]) 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 [546] which exacerbates predispositions to stress and anxiety [559]. 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 determining 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 ability to work together and their satisfaction with teamwork.

Understanding the crowd's 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 of collaboration may provide insights into why specific 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.

This research focuses on the impact of participants' personalities on ad-hoc online teamwork that is crowd-sourced, brief, and under pressure. We use the Big-5 personality model [200], 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-5 model as it is most commonly used for personality analysis (e.g., [257, 366, 242]) and for artificial intelligence systems that automatically adapt to personality (see [531] for a review of personality models used for personalization in persuasive technology, intelligent tutoring systems and recommender systems).

Table 5.1: Positive and negative facets of the BIG-5 personality traits [414]

Big-5 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

Many validated instruments exist to measure the Big-5 traits, including the brief version of the Big-5 Personality Inventory [476], which we use in this paper. The Big-5 model distinguishes between 5 traits[1], each of which has multiple facets (see Table 5.1)

5

5.2.2. 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 researchers improve 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 chapter is organized as follows. Section 5.3 presents and discusses related work, including an overview of traditional teams under pressure and crowdsourcing efforts in this domain and the study hypotheses. Section 5.4 describes the study design, including participant sample and task design. Section 5.5 describes the metrics used to capture participants' demographic characteristics, Big-5 personality traits, and ability (prior experience and self-perceived ability), as well as the metrics of teamwork, namely collaboration quality and communication patterns. Section 5.6 presents the results. In Section 5.7, we discuss the implications of this work, its limitations, and possible extensions for the future. Finally, section 5.8 concludes the paper with critical findings and closing remarks.

[1] Emotional Stability is often replaced in literature by its opposite Neuroticism.

5.3. Related Work

5.3.1. Teams in classical high-demand, time-pressing settings

Operational setting and problem scope

Significant research has been placed over the years on teams that need to perform in situations requiring spontaneous, ad-hoc decisions and short-term planning to resolve ambiguous or uncertain events and where the consequences of failure are significant [481]. The scope of the problems 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 [298], such as terrorism events [348], events occurring in long-duration spaceflights [500], nuclear plant control rooms [535], or situations taking place in a military context [149]. 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 [48], journalist teams for the immediate coverage of unexpected events [25], but also short-term project teams [182] and task forces [217]. Their size can vary, from dyads and triads [176] to dozens [235], to twenty or more [544].

Differences from regular teams

What separates these teams from teams in “normal” settings is the extreme, atypical environment within which they operate, which overall entrails time pressure, high levels of risk, increased consequences for poor performance [150], no previous work experience with one another, and the need to perform their task almost immediately on team formation [385, 392]. Harrison and Connors [226] use the term exotic environment to describe a work setting marked by hostile environmental demands, restricted working conditions, isolation from outside the setting, and confinement and enforced interactions for those inside it. Using the related term extreme environment, Bell et al. [46] 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 [185]. The tasks teams in these settings must solve usually characterized by ambiguity and urgency [535, 622].

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. Therefore, 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 [12]. Examining command and control teams, Ellis et al. [156] 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 the knowledge competencies of the team member with the most critical position 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 STAts (swift-starting action teams), is that they comprise experts [385] with specific roles and responsibilities. Multiple studies confirm the value of stable role structure in the division of labour and in enhancing the predictability of team interactions, allowing each team member to know what to expect from their teammates in critical situations [218, 535]. 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 [470, 597]. The meta-analysis of De Wit et al. [132] 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 benefit [69, 392]. A highly defined role structure with clear roles seems to help more structured tasks. On the contrary, a flatter structure may be better for ambiguous functions for which no apparent solution can be easily found [612] (such as the task of responding to the 2001 World Trade Center attack [392]).

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 [58]. Using the occupational personality questionnaire to study the emergency command ability of offshore installation managers, Flin and Slaven [170] 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, friendly, and humorous (high score on outgoing), (c) are less interested in analyzing human behaviour (low score on behavioural), (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's [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 among emergency nurses. Extraversion was also present among emergency department senior medical staff [62] as part of the controversial ENTJ (Extrovert, Intuitive, Thinking, Judging) personality type [2] [412].

[2]Whilst studies have been conducted on construct MBTI validity and test-retest reliability (including a meta-

Partially supporting these findings is the work of Wagner et al. [590] 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 [438]. By broadening the sample to the general public (virtual crowd), we aim to decouple the effects a specialized profession could have on one's propensity to emergency response.

Finally, certain *interaction patterns* are helpful 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 exact initial interactions affect subsequent communication and norms [192]. The study of Zijlstra et al. [633] reveals that specific early communication patterns 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 towards the task goal. Trust as a crucial factor is also highlighted [607]. The study of Waller et al. [593] reveals that efficient teams in non-routine situations focused their actions on information collection and task prioritization. Finally, Kanki and colleagues [279, 278] complement the above by showing that the communication of effective swift-start two-person crews focuses on immediate task execution, expressed as low complexity and 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 identify predictors of successful team formation in online crowd-action teams. Given that in a crowd setting, the allocation of roles is likely to occur based on arrival and availability, in this work, we focus on the parameters of

study by [77] which showed promising results), others have argued that there are scientific limitations to these studies, the use of MBTI, and its underlying theory (e.g., [63, 538, 457]).

personality and communication patterns as predictors of forming a successful crowd team to tackle unforeseen situations under time pressure.

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 made national societies unify and strengthen their approaches to natural, intentional, and accidental disasters [239]. 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 [401]. As emergency response evolves, 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 [68], war conflicts [1], and epidemics [325]. With the broadening digitization of services, society increasingly relies on technology., more intelligence, and a vast market of the Internet of things, software, and the worldwide web to enable widespread financial and data transactions [537]. Technological dependency makes us faster, more intelligent, and 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 [530], robotics [285], simulations [297], smart sensors [4], and social media networks [465]) to strengthen their outreach and preparedness.

Most emergency response teams operate hybrid, combining onsite support with online offsite communication. Some others divide efforts between online and face-to-face tasks depending on the response phase (i.e., mitigation, preparedness, response, and recovery [68]). Relevant to our research is the 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 first 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 [324], since their capabilities and duties differ, some their attributes are more imperative than others. The following subsections introduce two representative attributes critical for each teamwork domain.

Onsite emergency response teams. Two prominent attributes of onsite teams are **experience** and **coordination**. Teams working onsite are usually part of rescue operations [95] and disaster relief [50] 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 [616]. 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 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 emergency response. In their work on coordination in emergency response management, [97] developed a life-cycle approach with three distinct sets of activities on the timeline continuum (pre-incident, during incident, and recovery phases). The cycle closes after de-briefing and when actionable items are learned from the intervention and incorporated into the plan to affect future preparedness [97]. 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.

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 reach and distribute resources. Known crises overseen by offsite emergency response teams are air-traffic control [252], subway crisis management [231], and emergency response call centres [453, 425]. Although clear roles are essential in these teams, clear communication is of the essence. Depending on the kind of interaction (e.g., serendipitous, inbound, and outbound [318]) and the referent (e.g., non-experts' communication, situation update, situational awareness, services access assistance [582]), 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 prevent an escalation of the emergency [318]. Sensemaking has properties such as identity construction, retrospection, enactment, social reactions, dynamism, environmental cues, and plausibility [406]. 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 [597], most shortcomings from failed emergency responses are due to a deficiency in sensemaking (or contextual rationality). Weick's [1993] 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 [597] analyses the dynamics of role structure and sense-making in the Mann Gluch disaster. The incident served as an example of what needs to be reexamined about temporary systems, structuration, non-disclosed intimacy, inter-group dynamics, and team building [597], 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 [392]. These cognitive factors, defined by Mendonça [392], 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 [510] 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 [364]. In the following section, we explore crowdsourcing for emergency response.

5.3.2. Crowdsourcing for emergency response

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 primarily of individual, one-time, and remote interactions. This "long-tail" of contributions is a well-observed phenomenon in most content-oriented online communities [524]. 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.) [623, 233, 88], protest movements [158], but also activism [327, 163] and civic participation [236, 398]. In critical scenarios, the crowd is intended as a manifold social tool by serving 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. [249] 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. [533] proposes harnessing the self-organizing operation mechanisms of crowdsourcing for efficient disaster governance. In the context of natural disaster management, Ernst et al. [160] propose hybrid systems that rely on the remote coordination of volunteers to collect location-dependent information, which can support emergency managers making quick but solid decisions. Elsafoury [158] 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 to support communication between crowd users and disaster management systems by allowing seamless interaction. The platform "Ushahidi" [430] 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 under 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 [460]. 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 [15], Sahana [80], OpenIR [151], and CrisisTracker [487], provide support for mitigation and crisis preparedness. These tools pivot around the crowd to achieve 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 [621], or unfitness of the tools [142], 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 [232]. 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.

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 [451] and better coordinate their efforts [375]. Song et al. [533] analyzed twelve international crowdsourcing and natural disaster governance case studies. 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 [533]. 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 are building blocks for the entire social entity, 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 as in traditional face-to-face cases. Following Van de Ven et al. [577] 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 who are part of a virtual crowd and are often sporadic contributors. Therefore, while familiarity in crowd teams has tangible benefits [502] 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 jobs.

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. Lykourentzou 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 Lykourentzou 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 [585, 160, 171] 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 [112].

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.

5.4. 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 affect team success and perceptions of collaboration quality. Given the many facets and output measures considered, the study was exploratory to gain initial insights into what matters and how to be tested further in follow-up studies.

5.4.1. Sample

120 Amazon Mechanical Turk workers (41 female, 78 male, 1 prefer not to say) participated. The task duration was approximately 20 minutes. Most participants were of U.S. (67 users) and Indian nationalities (51 users); one was Irish, and another 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 5.7.

5.4.2. 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 of AMT workers as reported in 2018[225]. In selecting the payment amount, we considered three considerations from the literature[353, 433]. 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 who would only choose the task purely for its high compensation. Second, this payment had to cover the task duration. Thirdly, it considered the demographics of the target worker population (minimum wage).

We recruited through the Amazon Mechanical Turk (AMT) Human Intelligent Task (HIT) platform [3]. AMT was chosen for its breadth of crowd workers and its abundant labour availability, estimated to be no less than 2K workers at any given time and over 100K workers overall [139].

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 5.7.2. Finally, the HIT contained information about the reward, the duration of the task, and a short description of the cooperative game.

[3] AMT worker's population is composed primarily of Indian and American nationalities, followed by Chinese, British, and Philippino [139]. The gender is slightly predominantly female within the American sample and more male in other countries [139]. Its population average age is less than the world population average, as most AMT workers were born after the 1990's [139].

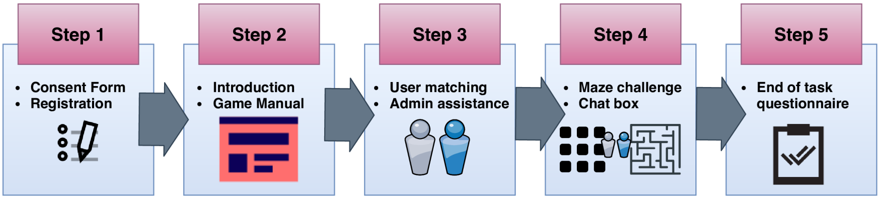


Figure 5.1: System overview with the five steps of the study design. After registration, users arrive at an introductory page with relevant information about the task, and then they are matched in dyads on a first-in-first-out basis. Each team then proceeds to their dedicated virtual room, cooperating to defuse the bomb in the maze within a given time frame. Finally, they completed a questionnaire about their abilities and perceived collaboration quality.

5.4.3. Task Design and Setting

Although the task was artificial, it was designed as an analogue setting, enacting the key characteristics of the high-demand, high-pressure environments 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 were playing a game, the element of physical danger, even an enacted one, alters their perception, possibly affecting how they process information, coordinate their efforts, and discuss[276].

2. **Pre-determined team roles**.

   The presence of these roles enables stable and predictable group interactions[388] instead of relying upon the slower and autonomous differentiation of team roles [45], which cannot always happen in circumstances of emergency. Predefined role-playing exercised control over one's limited access to information, symbolizing the relationship between an overseeing entity (in our case, the Lead Expert) and an operative agent (in our case, the Defuser). Furthermore, like real-life action teams, team membership symbolizes work shifts [633]. 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[315, 573].

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, users must act decisively within tight time schedules, often only with access to incomplete or difficult-to-decode information [82]. It means that the participants a) absorb information rapidly, b) judge by doing, c) decide on the spot, and 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)[150]. The combination

5

of elements, namely, high-stake, time-constrained, fractional information, and role inter-dependency, makes this particular task reasonably stressful. More so, the original game 'Keep Talking Nobody Explodes' has been used as a tool by past research to assess the effects of realistic stress on behavioural and physiological responses of participants [326, 495].

These studies confirm that controlled environments 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 5.1, 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-5 personality traits (Table5.3). By registering, the participants agreed with the terms of service and gave their informed consent.

Step 2: Introduction and game instructions. After logging in to the "dangerous and challenging world of bomb defusing"[302], the introductory page offered example screenshots of the two roles, instructions about the gameplay, and 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 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. Figure5.2 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. Figure5.3 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 (grey squares) and the path to the bomb (white sections). The Lead Expert could neither see the Defuser in the maze nor the bomb. The Lead Expert and Defuser could see the

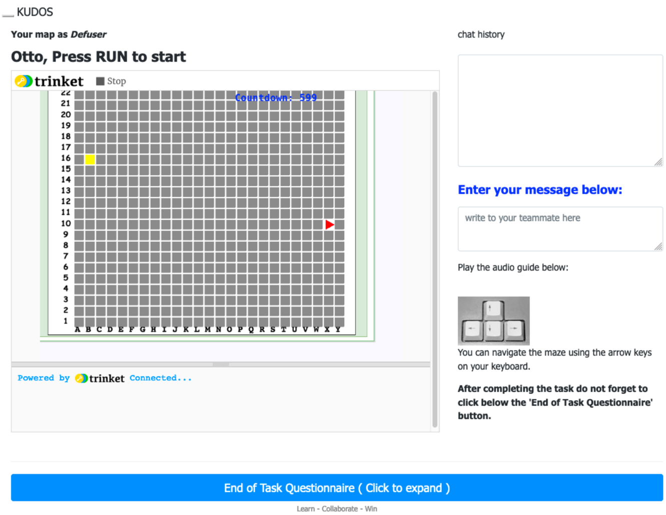


Figure 5.2: Defuser's view of the maze. The maze did not indicate the path to the bomb (red triangle), nor the walls. The participant was prompted to get directions from the Lead Expert through a chatbox (top-right of the screen).

same countdown and Cartesian coordinates of the maze, the chatbox, and the final questionnaire link.

The video game inspired the Maze module Keep Talking Nobody Explodes [302]. It consisted of a 25x25 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 maze information; they needed to cooperate. The Defusers could move inside the maze using 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 seconds. After finishing the game, the participants received a validation code to submit to the AMT HIT to get 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

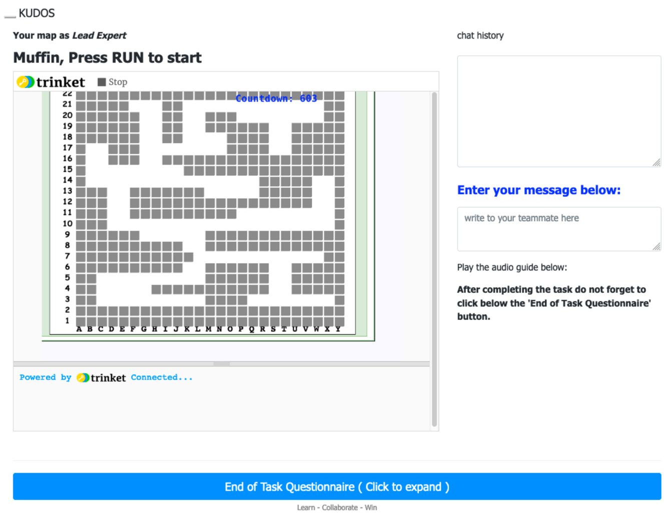


Figure 5.3: Lead Expert's view of the maze. The participant could see the map but did not know where the bomb and the Defuser were placed on the map.

added to the task complexity[4].

Hence, we did not include penalties for the Defuser colliding with a wall. The only liability – and the end of the 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 5.7.2.

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

5.5. Metrics

We grouped the multilevel approach into two classes referring to input and output variables (Table 5.2 summarises all variables, their type and range.). Here, the input metrics serve as the independent and output variables as the dependent variables.

5.5.1. Input variables

Big-5 personality traits

To measure the Big-5 traits within the context of large-scale assessment under limited time and resources, we used the Big-5 Inventory-10 (BFI-10) [476][7]. The inventory

[4]Also requiring considerably longer instructions and the introduction of manipulation checks to ensure instructions were read, which further adds to task complexity

[7]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-5 instruments, its

Table 5.2: Summary of variables

		Variable	Type	Range
Input	Personality [5]	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
		Mean	Interval	2-10
	Demographics	Gender	Nominal	Male, Female, Other, not-disclosed
		Age group	Ordinal	<20, 20-29, 30-39, 40-49, 50+
		Nationality[6]	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

Table 5.3: $BFI - 10$ instrument used, and it is scoring: the trait for which each item was used and whether it was reverse scored (R). Reverse score means that one is changed into 5, 2 into 4, 4 into 2, and 5 into 1.

BFI-10 Instrument						Scoring	
I see myself as someone who …	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly	Trait	Reverted
1. … is reserved	(1)	(2)	(3)	(4)	(5)	Extra.	R
2. … is generally trusting	(1)	(2)	(3)	(4)	(5)	Agree.	
3. … tends to be lazy	(1)	(2)	(3)	(4)	(5)	Cons.	R
4. … is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)	Neuro.	R
5. … has few artistic interests	(1)	(2)	(3)	(4)	(5)	Open.	R
6. … is outgoing, sociable	(1)	(2)	(3)	(4)	(5)	Extra.	
7. … tends to find faults with others	(1)	(2)	(3)	(4)	(5)	Agree.	R
8. … does a thorough job	(1)	(2)	(3)	(4)	(5)	Cons.	
9. … gets nervous easily	(1)	(2)	(3)	(4)	(5)	Neuro.	
10. .. has an active imagination	(1)	(2)	(3)	(4)	(5)	Open.	

contains ten questions (see Table 5.3). Derived from the shortening of its lengthier predecessor (the Big-5 Inventory (BFI-44)[476]), it focuses on the psychometric characteristics of the BFI-44's most representative items. It reduces each Big-5 dimension to 2 BFI items. The BFI-10 measures the personality traits of Extraversion, Agreeableness, Conscientiousness, Emotional Stability (Neuroticism), and Openness to experience [476]. 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 the mapping of statements to traits and which statements' scores are reversed in Table 5.3)[8].

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: the aggregation of group members' preferences [379] and uses aggregation strategies from Social Choice Theory [518]. [519] 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, most 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 [9,10]. Minimum is used as

correlations with self-and peer-ratings, and its associations with sociodemographic variables suggest good validity of the BFI-10 inventory despite its brevity [476].

[8]Reversed means that a score of 1 is changed into 5, 2 into 4, 4 into 2, and 5 into 1.

[9]which in the Group Recommender community are called respectively Least Misery and Most Pleasure

[10]Personality traits likely differ on whether a high (or low) trait level positively or negatively impacts team performance. Using minimum and maximum ensures this is no longer an issue.

one may expect that team performance is strongly affected by the weakest member in the team, in line with the famous 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 [428]), as the literature indicates the impact of diversity within teams. [11]

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 behaviour, culture, and experiences [322]. These socio-demographic factors provide further insight into the composition of teams and what other characteristics – aside from personality traits – influence 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 dispersed teams [405] geographically.

Multiple studies [431, 494] 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 [494]. Other demographic differences, such as race, sex, age, and nationality, have also been found [377] to affect the collective creativity of virtual teams. Age differences condition the creative processes of teams and intensify differences in technical experience [377]. Differences in nationality have a negative effect by interacting – however indirectly – with the technical experience of the teammates [377].

Communication patterns

The methodology by Bowers et al. [61] introduced a new approach to communication analysis prompted by a prior research gap in metrics that missed analyzing 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) **acknowledgements**, which were one-bit statements following the uncertainty of action statements, such as yes, no, roger; (d) **responses**, which differed from acknowledgements 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 crews' performance during simulated flight tasks, which improved the make-up of communication sequences analysis.

Based on Bowers et al. [61] contribution, Davaslioglu et al. [125] developed the Collective Allostatic Load Measurers system (CALM), which collected, aggregated, and analyzed

[11] 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 [428] and for generalizability to larger groups.

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.

Acknowledgement statements, for instance, predominated more amongst high-performing teams, while low-performing teams had higher portions of non-task-related statements. Similar studies on team communication analysis [633, 455] have identified communication patterns. Given the proximity of our methodology to the studies of Bowers et al. [61] and Davaslioglu et al. [125], 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 team members are taking action inside the game or 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 towards achieving a common goal. Planning statements 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 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 categorized as non-related when they do not contribute to achieving the goal (e.g. "What is the weather like?").

Table 5.4 illustrates an extract of the annotated chat between the Lead Expert and the Defuser. The patterns were labelled for each participant's text entry and annotated by two independent evaluators. The inter-rater agreement of the annotation was sufficiently high to be used in the study (Cohen's κ = .998, p = .000). In addition to counting how often each communication category was used, we also calculated the total number of posts made (chat total) and the number of words used (chat length).

5.5.2. Output variables

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 out-

Table 5.4: 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

puts), the expectations of its members, or its cost and time goals [20]. 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.

Perceived Collaboration quality

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

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 individual and team-level task-work and teamwork processes [308].

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 [44], 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 [55]. More so, the meta-analysis by Beal et al. [44] 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.

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 about how a team should function. It can also help disclose how the respective individuals communicate with the other team members and how they perceive the communication ties [111].

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 [596].

Perceived Balance. The metric addresses the question: *"Did both team members 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 are distributed between teammates and how they get carried out against the team's objectives [578].

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 not a sufficient guarantee that the teammates' behaviour aligned with the given role. By assessing the aspect of perceived balance through the lenses of the teammates, we could better understand the participants and whether it was a balanced act or whether a role was considered more demanding and accountable for the outcome than the other.

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.

5.6. Results

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

1. **Team performance:**
    - **Section 5.6.1** analyzes the effect of personality at team level [12], 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 5.6.2** analyzes the communication patterns using a one-way ANOVA to compare winners and losers, but also to compare the differences in behaviour between the team roles.
    - **Section 5.6.3** evaluates the impact on team performance of the participants' socio-demographic characteristics, using Chi-square tests and regression analysis.
2. **Perceived collaboration quality:**
    - **Section 5.6.4** assesses the relationship between personality traits and perception of collaboration quality, using correlation analysis for the individual traits.

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

- **Section 5.6.5** assesses the relationship between personality traits and perception of collaboration quality, using correlation analysis for the team traits.
- **Section 5.6.6** examines whether individual demographic characteristics played any role in people's perception of their collaboration, using one-way ANOVAs.
- **Section 5.6.7** analyzes the relationship between the communication patterns and the collaboration quality metrics, using correlation analysis to consider the roles of the Defuser and Lead Expert.

Given the many factors considered (e.g., five personality traits with four 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[13] to avoid Experiment Type I errors would reduce the power of the statistical tests to such an extent that Type II errors would be highly likely. Few insights would be gained[14].

We have, therefore, not applied such corrections (except in post-hoc pairwise comparisons). The study is exploratory, and the statistical results presented provide initial insights that lead to hypotheses for follow-up studies.

5.6.1. 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 5.5 shows these four metrics' mean (and standard deviation) for the winning and the losing teams. In winning teams, minimum Openness was significantly better (Mann-Whitney U=485, p=.02). There were no other significant results[15]. A binary logistic regression with the minimum metric[16] considered the effects of the team's personality on the likelihood of winning[17].

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=.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

[13] Less conservative corrections such as Tukey are not possible due to the data often not meeting normality assumptions

[14] Additionally, as many measures were not independent, Bonferroni corrections would also have been less appropriate

[15] Including no impact of Neuroticism or differences of standard deviation

[16] 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.

[17] Hosmer and Lemeshow test were not significant, thus, the model assumptions were met.

Table 5.5: Mean (Stdev) of standard deviation, average, minimum, and maximum for the Big-5 personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) for winning and losing teams.

		Open.	**Con.**	**Extra.**	**Agree.**	**Neuro.**
	StDev	1.06 (0.68)	1.41 (1.46)	1.15 (1.36)	1.50 (1.59)	1.10 (1.00)
Winning	Average	8.13 (1.51)	7.75 (1.53)	5.75 (2.32)	6.94 (1.53)	4.22 (2.33)
Teams	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)
	StDev	1.72 (1.36)	1.11 (1.32)	1.66 (1.52)	1.46 (1.25)	1.96 (1.88)
Losing	Average	7.26 (1.60)	8.24 (1.41)	5.01 (1.55)	6.40 (1.26)	3.82 (1.74)
Teams	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)

5

with an increased likelihood of winning (Openness: Exp(B)=1.52, Wald[18]=4.61, p=.032; Neuroticism: Exp(B)=1.58, 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 critical 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. The interdependence between roles reasonably determines this.

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

Table 5.6 shows the number of posts per chat category for winners and losers, winning and losing teams, and 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 of 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 ($F_{\text{action}}(1,118) = 4.426$, p= .038, $F_{\text{response}}(1,118) = 4.983$, p= .027).

[18]Wald is basically t^2 which is χ^2 distributed with df=1, and is used with small sample sizes instead of t.

Table 5.6: 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)

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 (Exp(B)=1.28, Wald=5.35, p=0.021; Exp(B)=1.21, Wald=3.92, p=0.048 respectively).

5

Increasing the chat length was associated with a decreased likelihood of winning (Exp(B)=0.97, Wald=4.04, p=0.044). These results 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 insufficient for success in online emergency response team settings since we could not find a correlation or causality between these variables.

Lead Experts used the Action category significantly more than Defusers ($F_{\text{action}}(1,118)$ = 14.736, p < .001) whilst Defusers used the Factual category significantly more F_{factual} (1, 118) = 5.273, *p*= .023). The Lead Experts have 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% versus 33%, *p*=.043) and a lower proportion of Uncertainty messages (8% versus 22%, *p*=0.041).

5.6.3. Impact of socio-demographic characteristics on performance

Table 5.7 shows the demographics of winners versus losers, excluding cases with very low frequency[19]. Pearson Chi-square tests show a significant association between gender and winning (χ^2(1, N=119) = 4.78, *p*=.029) and age and winning (χ^2(3, N=120) = 8.09,

[19] Namely prefer not to say for gender, and British and Irish for nationality, all with frequency 1

Table 5.7: Demographics overall and of winners versus 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	wom.	same	diff.	USA	India	same	diff.	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
win.	33%	15%	30%	22%	19%	35%	27%	26%	22%	36%	27%	0%	11%	33%	28%	27%
lose.	67%	85%	70%	78%	81%	65%	73%	74%	78%	64%	73%	100%	89%	67%	72%	73%

p=.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 (χ^2(1)=5.12, 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 5.6.1) was statistically significant (χ^2(3)=27.97, p <.001). The model explained 31% (Nagelkerke R^2) of the variance in winning whilst correctly classifying 78.2%

Being female was associated with a decreased likelihood of winning (Exp(B)=-1.31, Wald=4.97, p=.026). Similar to our earlier results, increases in minimum Openness and Neuroticism were associated with an increased likelihood of winning (Exp(B)=.47, Wald=11.92, p=.001; Exp(B)=.52, Wald=11.94, p=.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 less than personality. When age, nationality, or education are added to the binary logistic model instead of gender, they are insignificant.

5.6.4. 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. Overall, all perceived collaboration metrics were positively correlated (see Table 5.9) and for winners.

In contrast, for losers, the correlations with Satisfaction were insignificant (see Table 5.9), 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.

Table 5.8: Correlations between perceived collaboration quality metrics and personality traits, **=p<.01, *=p<.05

		OPEN	CONS	EXTRO	AGR	NEUR
All (N=44)	Performance	.062	-.187	.044	.434**	.106
	Cohesion	.050	-.181	-.088	.319*	.160
	Communication	-.111	-.256	-.217	.221	.159
	Balance	-.029	-.203	-.196	.317*	.318*
	Satisfaction	-.003	-.035	-.074	.032	-.031
Winners (N=24)	Performance	.081	-.099	.064	.289	-.023
	Cohesion	.053	-.148	-.006	.241	.013
	Communication	-.068	-.098	-.239	-.074	.044
	Balance	-.319	-.302	-.345	.354	.285
	Satisfaction	-.086	.144	-.009	-.072	-.098
Losers (N=20)	Performance	.013	-.336	.017	.761**	.330
	Cohesion	.021	-.226	-.162	.456*	.388
	Communication	-.178	-.551*	-.159	.547*	.397
	Balance	.315	-.053	.004	.338	.361
	Satisfaction	.025	-.233	-.112	.242	.050

Neuroticism significantly correlated with only Balance (see Table 5.8). Considering only winners, no significant correlations existed between the personality traits and any metric. In contrast, losers positively correlated significantly with 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 team's performance, communication, and cohesion[20][21].

We do not know whether being more conscientious made losers feel worse about their team's communication or whether their Conscientiousness influenced the team's communication. The lack of a significant correlation for winners points towards the first explanation, with Conscientious people perhaps being more honest in assessing team communication quality.

[20] 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.

[21] This seems more likely that Agreeableness influenced the performance, communication, and cohesion itself, indeed, given the lack of correlations for winners.

Table 5.9: Spearman correlations between perceived collaboration quality metrics, **=p<.01, *=p<.05

		Performance	Cohesion	Communication	Balance	Satisfaction
All (N=44)	Performance	1	.751**	.593**	.449**	.525**
	Cohesion	.751**	1	.649**	.528**	.502**
	Communication	.593**	.649**	1	.506**	.508**
	Balance	.449**	.528**	.506**	1	.389**
	Satisfaction	.525**	.502**	.508**	.398**	1
Winners (N=24)	Performance	1	.732**	.648**	.486*	.568**
	Cohesion	.732**	1	.725**	.512*	.579**
	Communication	.648**	.725**	1	.530**	.646**
	Balance	.486*	.512*	.530**	1	.484*
	Satisfaction	.568**	.579**	.646**	.484*	1
Losers (N=20)	Performance	1	.734**	.523*	.302	.299
	Cohesion	.734**	1	.514*	.419	.319
	Communication	.523*	.514*	1	.470*	.283
	Balance	.302	.419	.470*	1	.261
	Satisfaction	.299	.319	.283	.261	1

5.6.5. 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 team's perceived collaboration quality metrics by taking the average of its members, or only one member had provided their ratings by using that member's ratings. Average and minimum Openness positively correlated with perceived performance[22] 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[23], in line with our earlier observations regarding the impact of Agreeableness on individuals' opinions. The most interesting result is the significant positive correlation of all perceived quality metrics with Conscientiousness standard deviation[24][25]. A lower Conscientiousness standard deviation correlated with negative team feelings.

In a dyad, the lowest Conscientiousness standard deviation is when two people who are very similar in Conscientiousness work together—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 will likely yield good performance. It seems that the best performance- from the team member's point of view- for this particular type of task comes from two people working together who differ in Conscientiousness.

[22] Spearman correlations average Openness: r=.398, p=.02; minimum Openness r=.410, p=.02
[23] Spearman correlation: r=.400, p=.02
[24] Spearman correlations Performance: r=.644, p<.0001; Communication quality: r=.492, p=.003; Cohesion r=.403, p=.02; Balance: r=.448, p=.008; Satisfaction: r=.417, p=.01.
[25] There was also a significant Spearman correlation for minimum Conscientiousness: r=-.423, p=.01

Table 5.10: Mean (standard deviation) of collaboration quality metrics by gender and age, and also for teams that include the same or different genders

	Gender				Age			
Collaboration	**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)

Table 5.11: Mean (standard deviation) of collaboration quality metrics by nationality and education level, and also for teams that include the same or different nationalities

	Nationality				Education Level			
Collab. Metrics	**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)

5

5.6.6. Impact of socio-demographic characteristics on perceived collaboration quality: no significant result

Tables 5.10 and 5.11 show the perceived collaboration quality metrics for the different genders, age groups, nationalities, and education levels. One-way ANOVAs showed no significant effect of socio-demographic variables on perceived team performance, cohesion, communication, balance, and satisfaction[26]. 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). However, this was not statistically significant, which is unsurprising given the high variance and the sample size.

5.6.7. Impact of communication patterns on perceived collaboration quality: positive correlation

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[27]).

Satisfaction was positively correlated with the *Factual* category (r=.308, p=.042, for both the individual and team), also for Defusers (r=.457, p=.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=.437, p=.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

[26]There was a significant difference for education level on balance, but given the small numbers in all groups

[27]Given the low number of teams where both members responded, we used the perceived collaboration quality at the individual level only.

(r=.504, p=.020, for the team). The more cues were shared among the team members, the better Defusers perceived the team performance.

Balance was negatively correlated with the *Uncertainty* category (r=-.378, p=.011, for the individual), also for Lead Experts (r=-.440, p.=036; r=-.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=.457, p=.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 summarise, several communication categories correlate with perceived collaboration quality, with the role in the team impacting which categories matter. For a good perceived collaboration quality, it seemed necessary for Defusers to provide facts and for neither the team nor the Lead Expert to ask too many questions.

5.6.8. Post-hoc analysis on the 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 research shows that personality scales can be generalized across cultures [476, 489], the distribution in cultures of personality traits differs. Therefore, statine scores [566] are sometimes used for personality tests to normalize scores based on participants' country of origin.

5

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 5.12 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 Extraversion and Agreeableness.

In our sample, there were significant differences in Openness, Extraversion, and Agreeableness between the sample from India and the USA[28]. 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 were 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; for example, [71] 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 specific success rate on previous assignments). Similarly, one could imagine that coming from India and working on an American platform requires a certain level of Openness to Experience.

[28] Post-hoc test, Mann-Whitney U=811.5, U=611.0, U=933,5 respectively, with p<.001 (and still significant if Bonferroni corrected)

Table 5.12: Mean and standard deviation of the Big-5 personality traits in the literature [41] 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)
	India	5.16 (1.7)	5.52 (1.74)	5.18 (1.74)	4.19 (1.69)	5.41 (1.72)
Our sample	USA	6.69 (2.19)	8.34 (1.95)	4.13 (2.02)	5.85 (1.83)	5.88 (2.92)
	India	8.55 (1.56)	7.80 (1.89)	6.71 (1.89)	7.43 (1.74)	6.35 (2.02)

It may also impact whether people work with somebody from their own or another culture in the task. 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 5.7 and 5.11 respectively). There was no difference in 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[29].

5.7. Discussion, Limitations and Future Work

5

5.7.1. Discussion

In this chapter, 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, which 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 of Openness to experience (minimum).
- **Personality and perception**: Crowd workers' Agreeableness and Conscientiousness likely shaped their perception of the collaboration. Furthermore, dyads that

[29] Perceived performance was significant at p<.05, but not when Bonferroni correction was applied.

combined people differing in Conscientiousness were perceived by the participants to perform better.

- **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 these results and connect them with the broader teamwork literature in the coming sections.

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 [119, 39, 565, 104] 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 [434] 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 beneficial in teamwork[183], especially in situations of high stress, high stakes and limited time. Moreover, intellectual curiosity with regards to new circumstances is a characteristic observed in people with high Openness to experience [382]; this same trait is closely related to team creativity[507]. Substantiated by literature[507, 382], our results suggest that Openness may be more influential than task familiarity in determining the team's success.

Focused communication patterns get the teams going

From 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 communicated more effectively with their teammate and promptly devised 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 their paired teammates (Lead Experts) gave and responded over the chat by describing where they were at that point in the maze.

These results indicate the importance of *focused communication* (focusing on efficiency and action clarity), especially when the stakes are high and time-bound. Identifying collaboration patterns has also uncovered how winning individuals intervene during 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 techniques 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 [232], indicating that citizens are willing to become knowledgeable of emergency response measures and proactively contribute to community relations.

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 [65] and the trait faceted with trust, altruism, and humility [381]. As highly agreeable people tend to be more sympathetic toward others [564] and more humble, this may have made them more forgiving towards their teammates and themselves. We also found that individuals in heterogeneous teams 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 heterogeneous teams in Conscientiousness does not have to be detrimental to actual performance, as shown by our other results and Mohammed and Angell [400]. Our result conflicts with that of Gevers and AG Peeters [193], 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.

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 the team role. Defusers seemed more satisfied with the collaboration when they and the team used more Factual statements. Lead Experts seemed less pleased 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 seemed to have expected specific communication patterns from their teammates, and these greatly depended on what part of the information they had access to. Defining clear roles is essential, as team role clarity improves collaboration [26] and communication styles aligned with team roles matter for effective and satisfactory teamwork (as shown in this chapter, and line with [131]). It may be even more vital in high-pressure tasks with high interdependence.

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 [600].

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 5.7 for descriptives for same-gender teams and teams with different genders). Apesteguia et al. [23] 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 5.10 for descriptives). There was a significant impact only on Communication (post hoc, Mann Whitney U=268, p<.005, Bonferroni corrected), with Communication appreciated more in same-gender teams. As there is a big difference between India and the USA in gender equality (the USA is 30^{th} (out of 156) in the Global Gender Gap Index [175] compared to India only being 140^{th}), 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<.05, Bonferroni corrected) and similar trends for Communication and Performance (p=.1 after Bonferroni correction), with all being perceived better for same nationality teams.

We considered whether the impact of gender on winning might 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 [138].

5.7.2. Limitations

Exploratory study

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

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 quickly and based on availability. For this reason, it is 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 [260] would assist the matchmaking process according to pre-defined criteria. Other matching systems, such as AI (machine learning and feature extraction), could also be used as baselines.

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 behavioural, 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 based on their reputation and the 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 overlooked apparent communication issues due to the language in the chats. Nevertheless, future studies will include a test to ensure an appropriate English proficiency.

The absence of pre-screening in English also has a positive aspect, as it means our study can be generalized to emergency crises where English is not necessarily the native language whilst still being used for essential 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.

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. 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 emergencies to test the generalizability of our findings[30]. 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 time limit of 400 seconds 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 uncover differences in behaviour and collaboration that we could not identify by running a single study design.

In our experiments, the maze's walls were 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 assess their significance.

5.7.3. Implications and Future Work

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 [351, 427]. The task impacts what team attributes matter for good actual and perceived performance and collaboration. For the emergency task studied in this chapter, 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 the emergency response teams 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. Still, 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 [267]. More so, previous research on the effects of personality traits in teamwork did not consider the impact of the task type under stress [119, 39, 565, 104], 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 benefit the

[30]However, there are apparent ethical issues with this

crowdsourcing and online work fields that are becoming ever so relevant due to recent and significant changes in how 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 [372].

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, 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.

Conversational AI support for remote emergency response teams

The analysis of the communication patterns indicated that not all teams focused on task execution correctly since some adopted less-than-optimal communication strategies. Our results provide insights into which communication acts may be essential and 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 [43] (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 [7], most forms of crowdsourced self-organized teams (e.g., neighbourhood watch [32]) could themselves 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 like 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. Further, personality seems to have determined the propensity for more or less rigour and clarity in communication.

Considering the numerous variables and increased reliance on crowdsourcing for rescue operations and emergency response, we advocate for developing adaptive and personalized intelligent systems. AI-aided emergency response can provide support and knowledge to teams according to 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.

5.8. Conclusion

In this chapter, we present a study where 60 crowd dyads collaborated in a high-pressure, computer-mediated task and answer the Research Question **RQ3: How do personality and communication patterns affect online ad hoc teams under pressure in emergency response situations?** . The experimental design expected crowd workers to play complementary roles in a time-bounded critical scenario. We explored the possible impact of the participant's 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 better in executing this high-pressure task. The analysis of the team communication patterns suggests that teams communicating more through action-response loops were likelier 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' heterogeneity in conscientiousness seemed to make them feel better about 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. These experimental results show that the perception of 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's needs.

Our findings support future research on computer-based collaboration under pressure. It shows ways to tailor the development of AI to provide 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 conclusions presented here. In Chapter 6, we investigate how crowd workers form teams when given access to deep-level profiling information of a set of users. The study provides insight into decision-making in team formation. It proposes strategies to automate the team formation process based on crowd workers' decisions.

6

The Wisdom of the Crowd in Team Formation

6.1. Abstract

Team formation needs to consider crowd workers' profiling attributes, given their perceived usefulness (see Chapter 3) and their impact on team performance (see Chapter 5). Increasingly, crowdsourcing systems use automated team formation which automatically allocates crowd workers to teams. To make automated team formation more crowd-worker centered, we need to know how crowd workers think the profiling attributes ought to be used during team formation. This chapter therefore addresses the Research Question, **RQ4: How does the crowd decide on team formation given profiling attributes?**. Following a User-Centered approach, we asked 102 crowd workers to divide a list of individuals (with given attributes) into teams to understand their approach to the team formation problem. For the task, we chose an online education/school scenario since forming teams of learners is a relatable experience (i.e., most people have been part of a study team at some point in their education). Furthermore, forming teams of learners online shares common challenges with team formation with crowd workers, such as a lack of in-depth knowledge and familiarity with the individuals that must be matched into teams. In this work, we present a User as Wizard [380] study where participants were asked to form four teams of three teammates from a pool of twelve dummy learner profiles. The profiles contained information about the learners' Conscientiousness, Openness to experience, and (Cognitive) ability levels. These attributes came from a pre-study with a smaller sample of crowd participants (N=52) rating the relevance of the Big-5 personality traits and (Cognitive) ability for team formation for educational purposes.

6.2. Introduction

In this research, we approach the team formation problem through the eyes of the crowd and in a human-centred way, thus addressing the Research Question **RQ4: How does the crowd decide on team formation given profile attributes?**. In an online team formation scenario, we investigate crowd users interacting with a system as they assign students to teams online. Through this work, we assess what future automated systems should consider when recommending teammates and team compositions. Automated strategies for the Team Formation Problem (TFP) already offer solutions through computed outputs [427]. Some of the most common forms of computed solutions rest on established partitioning approaches (e.g., regression analysis optimization [407], genetic algorithms [66], k-means [17], etc.). However, algorithmic modelling does not always mirror the cognitive processes behind human decisions during team formation and may fail to capture the subtleties and richness of human decision-making. Factors such as an intuitive understanding of interpersonal dynamics, the evaluation of compatibility based on personality traits, and the subjective perception of individual ability levels can strongly influence the formation of teams. Our research seeks to explore these implications for automated team formation. *Our study examines what happens when users form teams manually without following a specific strategy, aiming to elicit human decision-making insights to guide automated team formation.*

6.2.1. Study Focus and Related Work

Our research aligns closely with studies by Odo [429] on team formation for collaborative learning, focusing on individual traits such as personality and ability levels. By gaining a deeper understanding of human decision-making processes, we aim to enhance automated team formation systems, making them more responsive to the nuances people consider when forming teams. This could lead to more effective, cohesive, and productive teams in various settings, including online learning and corporate environments. Specifically, we take into account Odo's [2021] salient conclusions that personality traits and specific metrics (i.e., *distribution of characteristics, team cohesion*, and *team balance*) significantly impact team formation. Furthermore, we consider Odo's 2021 findings on the relevance of Conscientiousness as a crucial attribute in team formation, especially in collaborative learning settings. Their study observed that users tended to distribute Conscientiousness more evenly than some other personality traits, such as Emotional Stability and Extraversion[1]. High Conscientiousness was also often more balanced in specific team sizes (i.e., 4 to 6 people) followed by the Ability levels of the teammates. Lastly, distributing attributes across teams to ensure cohesion (i.e., members of the same team sharing similar attributes) was a known strategy, especially for larger teams and with personality traits such as Conscientiousness and Emotional Stability. Building upon these findings on the importance of personality traits and evaluative metrics in team formation, we further Odo [429] research by focusing on a larger participant sample and using the personality traits Conscientiousness and Openness to Experience as well as Ability. Our study aims to broaden our knowledge of human decision-making processes of team formation in an online, remote context.

[1] Openness to Experience was not studied by Odo [429]

6.2.2. Research Questions and Hypotheses

Drawing inspiration from the approach adopted by Odo, we explore how participants incorporate chosen personality traits when forming learner teams. Our work enhances our comprehension of user-driven team formation. The study uses the User as Wizard (UAW) method, following the experimental protocol chosen for this study, aligning with Odo [429]'s approach. UAW, a method conceptualized by Masthoff [380], positions humans as the pivotal point in the design process, allowing participants to execute system tasks without requiring scripts or guidelines. Our investigation is focused on whether participants aim to distribute learner attributes or resources within and across teams to maintain a fair share of assets. Although there are many metrics for evaluating team formation and the distribution of resources within the team (e.g., team energy, roles, relationships, motivation, problem-solving, etc., [611]), fewer are the metrics for assessing teams whose members do not know each other or are not familiar with the task, for this research, we focus on a specific set of metrics drawn from the work of Odo [429], namely even distribution, cohesion, and balance. We also include team attributes' averages. With these considerations in mind, we formulated the following Research Questions:

- **RQ4.1: Does the even distribution of the learners' attributes differ based on the attribute (i.e., Openness to Experience, Conscientiousness, and Ability)?** This Research has a follow-up sub-question if the answer is true. The sub- question regarding potential disparities in attribute levels, namely high and low, is as follows:

  - **RQ4.1.1: Which attribute level is the most evenly distributed?**

  The sub-research Question concerns differences in even distribution between high and low attribute levels.

- **RQ4.2: Does cohesion differ based on the attribute?** This question focuses on the similarity of team attribute levels and whether teammates have similar characteristics.

- **RQ4.3: Does the team's balance differ based on the attribute?** This question investigates teams' high and low attribute levels, checking for a balance in the number of high and low levels for each attribute.

These questions shed light on the human processes involved when allocating learners into teams with limited information via computer-assisted systems. The subsequent parts of this chapter are organized as follows: Section 6.3 delves into the related work on forming learner teams and profiling attributes, including personality traits and cognitive abilities. Section 6.4 explains the choice of three profiling attributes for team formation, founded on a study involving online participants. Section 6.5 introduces the team formation tool designed for the UAW study, discussing the participants and the results of the team formation experiments. Section 6.6 analyses the findings. Section 6.7 discusses the limitations. Finally, Section 6.8 concludes the chapter.

6.3. Crowd teams in online education

The Team Formation Problem (TFP) is a usual concern for educators as they are in charge of classroom activities and must decide who should be teamed with whom. With insufficient resources such as narrow timelines, classroom size, and academic objectives (e.g., facilitating new collaborations between students, sharing ability levels across teams, etc.), educators face constraints limiting their investment in the TFP. Furthermore, in the conventional sense, forming teams can be considered a pen-and-paper problem. Flexibility and cost-effectiveness are generally two advantages of solving the TFP of learners manually. However, the growth of online classrooms (e.g., MOOCs – Massive Open Online Courses) and remote education have transformed team formation for learners, reducing it into an intractable problem when solved manually. More complications arise from the lack of time and familiarity with the students. Thus, many online tutors resort to letting online learners form teams alone or relying on tools that automate the TFP.

One advantage of automated tools for team formation of learners is the computerization of matchmaking. The algorithm in charge of the team composition treats attributes as variables and distributes them according to quantifiable objectives, such as an equal spread of academic grades across multiple teams. Many tools offer automated solutions to the TFP in education and use several criteria to profile learners [264]. The recent systematic literature review by Maqtary et al. [369] shows various team formation attributes and techniques within the educational domain to automate the team formation task. One example of such a system is CATME [527] and its Team-Maker tool that automatically forms teams based on student responses to various categories such as demographics, performance metrics, and convenience. In large-scale online education, research (e.g., [503, 602]) has experimented with criteria-based team formation algorithms, mainly yielding positive results. The systematic literature review by Odo et al. [427] on team formation for collaborative learning shows that, in research, no specific characteristics are considered when forming teams and no ideal team size. They also identify a gap in the literature regarding an analysis of team formation algorithms and their comparative performance in the collaborative learning domain.

6.3.1. The User as Wizard method in team formation and education

Most research on team formation in teaching and academic performance optimization relies on top-down algorithmic methods based on learners' modelling and predefined objectives. However, human-centred approaches to the TFP have been proposing the principle of co-design and user engagement in the system design process.

One approach that is sometimes used to facilitate studies of users interacting with computers is the Wizard of Oz method [211]. Conventionally, the technique combines two machines, one for the subject and one for the experimenter (the *Wizard* pretending to be a computer typing replies). One of the first implementations of The Wizard of Oz dates back to a study in 1985 [211]. The study featured the IBM Personal Computer used in several experiments with simulated user interfaces for an easy-to-use home computer banking program. Since then, the Wizard of Oz has been part of numerous other studies on human factors and human-computer interaction design.

In the TFP, the Wizard of Oz is mainly used to evaluate automated processes in co-operative scenarios such as human-robot-interaction [371, 569] and human-autonomy teaming [291, 389]. An alternative to the Wizard of Oz method is the User as Wizard method (UAW) formally introduced by Masthoff [380]. UAW predominately focuses on developing human-centred research to inspire algorithm adaptation. It places participants in the role of the Wizard and leaves them completely free to perform the task without a script to follow [380]. The method consists of two stages. One called *Exploration stage* sees participants taking the role of the adaptive system. The other, named *Consolidation stage*, requires participants to judge the performance of others. We focus on the exploration stage by presenting participants with a scenario (team formation in education) and fictitious users (dummy learners' profiles). The exploration phase has two main steps: 1) Giving participants the task the adaptive system is supposed to perform. 2) Finding out participants' reasons for their decisions and actions. In the TFP in education, the method is used in two studies by Odo [429] investigating automatic team formation. The first study used a combination of Conscientiousness, Agreeableness, and ability levels to characterize twelve learners. It then asked twenty-four participants to form different-sized teams to ensure they would work well together. The second study followed a similar procedure. However, it had sixteen participants and used Emotional Stability and Extraversion. The studies show that users account for personality and ability characteristics as they assemble teams of learners. Conscientiousness is weighted more than Emotional Stability, Agreeableness, and Ability in the distribution of traits. Another interesting was that team attributes such as cohesion and balance were considered as users formed teams of learners.

6

6.3.2. Learners profiling attributes

For the past 200 years, education was mainly mass schooling with little to no adjustment to the individual's characteristics [514]. However, a recent growing trend of personalized education meant that schools and universities are acquiring novel approaches to tailor education [551]. Personalized education systematically adapts instruction to individual learners [561]. The essential aspect of personalization – and subsequently ad hoc team formation with learners – is the capacity to gather information about each individual and classify it meaningfully. From collecting information about the student before the course starts to documenting their performance and engagement, modelling profiles can be a static procedure (one-off) or a dynamic process (ongoing). According to Drachsler and Kirschner [147], at least four characteristics differentiate learners: personal, academic, social/emotional, and cognitive.

Personal characteristics often relate to demographic information such as age, gender, language, socioeconomic status, cultural background, and specific needs (e.g., disabilities and impairments to learning). *Academic* characteristics are learning goals, knowledge, educational type, and educational level. *Social/emotional* characteristics deal with sociability, self-image (including self-efficacy and agency), mood, etc. Lastly, *cognitive characteristics* relate to attention, memory, cognitive flexibility, and cognitive skills.

Another important set of characteristics is personality traits. In the broadest sense, personality traits are individual differences that affect human behaviour [164]. Many personality models classify people based on their differences. One of the most used models in the education setting is the Five-Factor Model (FFM), also known as the Big-5 Model (OCEAN) [384]. In Chapter 3, we used this model when studying crowd teams in an emergency response task. Other known ones are the Dominance, Influence, Steadiness, and Conscientiousness model (DISC) [548] and the Myers-Briggs Type Indicator (MBTI) [383][2]. To measure various personality traits and models, there are several tailored instruments such as the HEXACO model of personality structure personality inventory [29], the Revised NEO Personality Inventory (NEO-PI-R) [114], the Eysenck personality inventory [485], the Minnesota multiphasic personality Inventory [73], the Birkman method [167], and many more. This chapter focuses on the Big-5 model and its five personality traits: Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

The Big-5 personality traits in education

The first version of the Big-5 came from Tupes and Christal in 1961. However, only in the 80s and 90s did it reach an academic audience through the work of Digman [141] and Goldberg [200]. Since then, the five major dimensions have provided a robust framework for many personality-centered studies. The Big-5 traits, or dimensions, are Openness to experience (inventiveness and curiosity; its opposite is consistency and caution), Conscientiousness (organization, efficiency, and responsibility; its opposite is extravagance and carelessness), Extroversion (assertiveness, sociability; its opposite is introversion), Agreeableness (compassion, friendliness, trust in others; its opposite is a criticism), and Neuroticism (tendencies toward sensitivity and anxiety; its opposite is confidence and resilience). Emotional Stability is also frequently used as a dimension indicating the opposite scale of Neuroticism.

6

In their meta-analysis of the Big-5 Model of personality and academic performance, Poropat [463] lists theory-grounded arguments justifying the relationship between personality and learner achievement across academic subjects. The first theoretical basis for the Big-5 is the hypothesis that behaviour and work outcome are related.

(Cognitive) ability in education

(Cognitive) ability is the collection of skills needed to complete tasks such as thinking, learning, reading, remembering, speaking, listening, and focusing; it is the capacity to think in the abstract, reason, problem-solve, and comprehend [459]. Over a century of scientific research has shown that general (Cognitive) ability (or *g*) predicts a broad spectrum of critical life outcomes, behaviours, and performances [311].

The educational setting often presents several domain-specific and general (Cognitive) ability tests [488, 361]. (Cognitive) ability instruments (e.g., the Miller Analogies Test [391]) are often present in educational admissions decisions as they estimate the relationship between (Cognitive) ability and performance. In the teamwork context for education, Liu et al. [345] proposes a two-stage framework to apply cognitive diagnosis

[2]Though there is some dispute about the validity of the model [429].

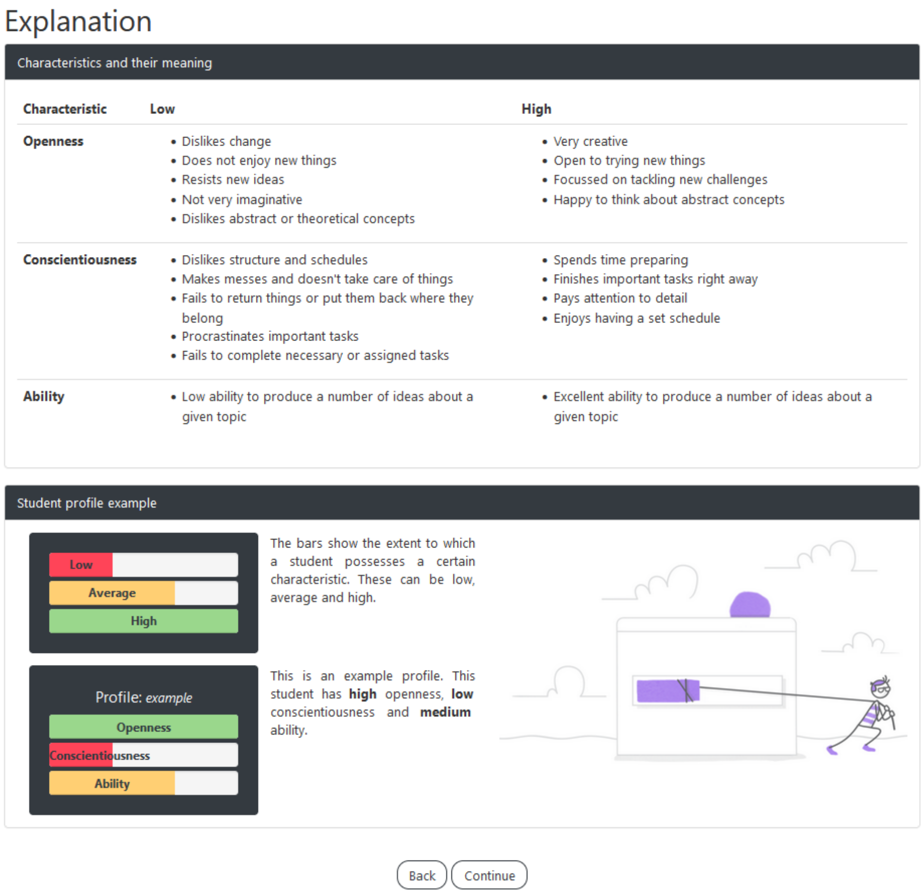
Table 6.1: Short descriptions of low and high profiling attributes inspired by Neuman et al. [414] as shown in the UAW study with crowd participants. The table did not include the medium range, as it was explained to be a central point between the two extremes.

Attribute	**Low**	**High**
Openness	Dislikes changes Does not enjoy new things Resists new ideas Not very imaginative Dislikes abstract or theoretical concepts	Very creative Open to trying new things Focused on tackling new challenges Happy to think about abstract concepts
Conscientiousness	Dislikes structure and schedules Makes mess and doesn't care about things Fails to return things or put them back where they belong Procrastinates important tasks Fails to complete necessary or assigned tasks	Spends time preparing Finishes important tasks right away Pays attention to detail Enjoys having a set schedule
Ability	Low ability to produce ideas Struggles with cognitive problems	Excels at producing ideas Excellent at solving cognitive problems

for collaborative learning team formation. One quantifies the student skill proficiency; the other optimizes team formation based on dissimilarity-based and gain-based objectives. More work using (Cognitive) ability as a modelling feature [9, 101, 362, 619] investigate the TFP as a simulation and do not formally test their approaches through real-world experimental studies. In this work, we compare (Cognitive) ability with personality traits (Big-5) by assessing how users assemble teams of learners manually given a set of dummy profiles.

6.4. Pre-study: Openness and Conscientiousness

Before running the UAW study, we carried out an exploratory study with a batch of participants (*N*=52) recruited from Prolific [447] to assess which of the Big-5 traits [492] users would consider most relevant for team formation in the education domain. The reason for looking at a sub-set of profiling attributes rather than using them all concurrently was to avoid excessive feature congestion [490] occurring when too many elements clutter the User Interface. We included (Cognitive) ability as a profiling characteristic known to affect performance [159] and considered in similar settings by Odo [429]. The results from the exploratory study allowed us to narrow down the attribute lists of the learners to a smaller subset (Openness, Conscientiousness, and Ability). In the survey, participants had to indicate on a Five-point Likert Scale (where 1=Not at all important and 5=Extremely Important) how much they perceived each personality trait necessary when forming teams of learners. The order of the attributes was shuffled to prevent presentation bias. Out of the Big-5 attributes, Conscientiousness (*mean*=4.05, *sd*=0.82), and Openness (*mean*=4.09, *sd*=0.99) were the top two preferred (Table 6.2).


Explanation

Characteristics and their meaning

Characteristic	Low	High
Openness	• Dislikes change • Does not enjoy new things • Resists new ideas • Not very imaginative • Dislikes abstract or theoretical concepts	• Very creative • Open to trying new things • Focussed on tackling new challenges • Happy to think about abstract concepts
Conscientiousness	• Dislikes structure and schedules • Makes messes and doesn't take care of things • Fails to return things or put them back where they belong • Procrastinates important tasks • Fails to complete necessary or assigned tasks	• Spends time preparing • Finishes important tasks right away • Pays attention to detail • Enjoys having a set schedule
Ability	• Low ability to produce a number of ideas about a given topic	• Excellent ability to produce a number of ideas about a given topic

Figure 6.1: Explanation page of the task. The first half showcases the Low and High characteristics of the three traits. The second half illustrates the learners' cards.

Figure 6.2: Overview of the team formation card-based drag-and-drop UI. It allowed users to form teams by placing learners into four separate containers representing four teams and to adjust the teams by dragging the learners' cards between them.

Table 6.2: Mean Importance, Standard Deviation (SD), and Standard Error (SE) of each Big-5 personality trait according to the pre-study participants (N=52). Their preference for profiling attributes for team formation of learners indicates that Openness and Conscientiousness are the most favoured traits.

Learners Attributes	*Mean*	*SD*	*SE*
Openness	4.09	0.99	0.13
Conscientiousness	4.05	0.82	0.11
Extraversion	3.30	0.94	0.13
Agreeableness	3.84	0.89	0.12
Neuroticism	2.61	1.25	0.17

In comparison, Agreeableness scored lower (*mean*=3.84, *sd*=0.89), followed by Extraversion (*mean*=3.30, *sd*=0.94), and Neuroticism (*mean*=2.61, *sd*=1.25). According to these findings, Openness and Conscientiousness are the most important personality traits when profiling learners for team formation. We used these attributes plus (Cognitive) ability (as it is typically another known attribute in education) to profile learners in the follow-up UAW study.

6.5. Main Study: Forming Teams

This section details our main study, including the team formation tool and procedure (Section 6.5.1), distribution of learners attributes and values across twelve dummy profiles (Section 6.5.2), metrics used (Sections 6.5.3), the demographics of the participant pool (Section 6.5.4), and the results (Section 6.5.5).

6.5.1. Team Formation Tool and Procedure

A web-based application was developed to facilitate the study. The application used Javascript for drag-and-drop functionality, HTML for rendering learner profiles, and Python Flask to bridge the gap between the front and back end. The application comprised a series of pages that guided participants through the study procedure, detailed as follows:

1. **Registration:** Participants were recruited from Prolific via a Human Intelligence Task (HIT). Through the HIT, participants could access the application URL, provide their consent to partake in the study, and then gain access to the task by setting up a username and password.

2. **Introduction:** The introduction page expressed gratitude for their participation, explained the scope of the task and outlined the next step.

3. **Explanation:** The Explanation page (Figure 6.1) presented a table outlining the characteristics that differentiate between high and low attribute levels (Openness to Experience, Conscientiousness, and Ability) in the learner profiles, adapted from Neuman et al. [414].

Table 6.3: Learners dummy profiles used for the UAW study with their profiling attributes scores (L=low, M=medium. H=high).

Learners	Open	Cons	Able
Andy	L	M	H
Bo	M	L	M
Carl	H	M	L
Darrel	M	M	L
Edwin	M	L	H
Finn	H	L	M
Grant	M	H	M
Hunter	L	M	M
Ian	M	M	H
Josh	L	H	M
Karter	H	M	M
Liam	M	H	L

Table 6.4: Research Questions (RQ4, RQ4.1, RQ4.2, RQ4.3) and Post-hoc Analysis mapped together with the study design metrics.

Research Questions	Metrics
RQ4: How does the crowd decide on team formation given profile attributes?	Summary of the following:
RQ4.1: Does the teams' even distribution differ based on the attribute (i.e., Openness to Experience, Conscientiousness, and Ability)? RQ4.1.1: Which attribute level is the most evenly distributed between teams?	Even Distribution
RQ4.2 Does the team's cohesion differ based on the attribute?	Cohesion
RQ4.3: Does the team's balance differ based on the chosen attribute?	Balance
Post-hoc Analysis	Team Average, Attribute Distribution Thematic Frequency (ADT)

4. **Team Formation Task:** Participants used the card-based interface to form four teams, each consisting of three learners. This was accomplished using the drag-and-drop functionality (see Figure 6.2).

5. **End-of-Task Questionnaire:** Participants answered the following question: *Explain your rationale behind teaming the learners as you did.* This open-ended question was where participants were encouraged to provide a detailed account of their thought processes and strategies when forming teams.

Table 6.5: Three examples of how attribute levels (3 high, 3 low, 6 medium) can be distributed between four teams. The first table shows the value count of High attributes to derive the Even Distribution (ED). The second table shows the coding of the values high=3, medium =2, low=1 to derive Cohesion. In the last table, the letters are used to indicate high (H), medium (M), and low (L) attribute values to calculate Balance.

	Even Distribution		
Team\Person	*P1*	*P2*	*P3*
T1	1	0	0
T2	0	1	0
T3	0	0	1
T4	0	0	0

	Cohesion		
Team/Person	*P1*	*P2*	*P3*
T1	3	2	1
T2	2	3	1
T3	2	1	3
T4	2	2	2

	Balance		
Team\Person	*P1*	*P2*	*P3*
T1	H	M	L
T2	M	H	L
T3	M	L	H
T4	M	M	M

6.5.2. Learners profiling attributes

We created twelve fictitious learner profiles (Table 6.3) comprising value-neutral culture names [219] and three profiling attributes (Conscientiousness, Openness, and Ability). Indicating differences between learners were three attribute scores shown as low (red, 1/3 of the progress bar), medium (yellow, 2/3 of the progress bar), and high (green, 3/3 of the progress bar). Participants were informed about the meaning of these scores in the explanation part of the study (Table 6.1, and Figure 6.1).

We distributed the attribute scores according to the following criteria: a) half of each attribute score (6/12) was medium, three were low (3/12), and the remaining three were high (3/12), b) no learner profile had more than one low and one high attribute score.

6

6.5.3. Metrics

To answer our Research Questions, we adopted the quantitative metrics proposed by Odo [429], namely *Even Distribution, Cohesion,* and *Balance*. We also present another set of metrics for the post-hoc analysis, namely *Team Average*, and *Attribute Distribution Thematic Frequency* (ADT). We explain how these metrics were calculated and how they map to the Research Questions (see summary in Table 6.3).

Even Distribution (ED). This study assesses whether crowd-formed teams have an even distribution of conscientiousness, openness to experience, and ability at both high and low levels. An Even Distribution (ED) can be described as an equal spread of attribute levels across teams. We use this metric to answer RQ4.1 (see Table 6.4). When calculating ED, we handle each attribute separately. Within our study design, the best ED occurs when three out of four teams have one high level of the same attribute. It is important to note that this distribution pattern for high Conscientiousness, high Openness to Experience, and high Ability also applies to other cases, such as low Conscientiousness, low Openness to Experience, and low Ability[3]. To measure this metric, we used the Even Distribution by Odo [429], which takes the standard deviation of the number of high attributes within each team. The calculation process is as follows.

[3]This limitation is determined by our study design where three attribute levels must be distributed between four teams, leaving at least one team without a high (or low) level of the given attribute

Table 6.6: All possible high or low attribute levels distributions between four teams. There can only be three high or low attributes.

	Distribution 1	**Distribution 2**	**Distribution 3**
T1	1 (high or low)	2 (high or low)	3 (high or low)
T2	1 (high or low)	1 (high or low)	0
T3	1 (high or low)	0	0
T4	0	0	0
Even Distribution (ED)	**0.43**	**0.83**	**1.30**
ED Class	Evenly dist.	Unevenly dist.	Unevenly dist.

Suppose we have four teams of three learners and want to assess the high Conscientiousness (Cons) ED. Let's say the teams have the following distribution of high attributes (illustrated in the first example from the left in Table 6.5).

- Team 1: 1 high Cons
- Team 2: 1 high Cons
- Team 3: 1 high Cons
- Team 4: 0 high Cons

The steps to calculate the standard deviation of this distribution would be:

1. Identify the number of high Cons in each team: $[1,1,1,0]$
2. Calculate the mean quantity of high Cons across all teams: $(1+1+1+0)/4 = 0.75$
3. Subtract the mean from each team's number of high Cons and square the result: $[(1-0.75)^2, (1-0.75)^2, (1-0.75)^2, (0-0.75)^2] = [0.0625, 0.0625, 0.0625, 0.5625]$
4. Sum the squared differences: $0.0625 + 0.0625 + 0.0625 + 0.5625 = 0.75$
5. Divide this sum by the total number of teams: $0.75/4 = 0.1875$
6. Extract the square root of the quotient from step 5 to get the standard deviation: $\sqrt{0.1875} \approx 0.43$

Therefore, in this example, the standard deviation for the high Conscientiousness distribution is approximately 0.43. The standard deviation can range from 0 to a maximum value depending on the attribute's distribution in the teams. A value of 0 would indicate that the attribute is perfectly evenly distributed among the teams, while a higher value would show a less-even distribution. In this context, a low standard deviation could be considered any value close to 0, while a high standard deviation would be significantly greater than 0, indicating a less-even attribute distribution. We perform this calculation for high Conscientiousness, high Openness to Experience, and high Ability, as well as for low Conscientiousness, low Openness to Experience, and low Ability. The calculation process is identical for low attributes, simply replacing high attributes with low ones. For instance, for low Conscientiousness, we would count the number of low Conscientiousness in each team and follow the same steps outlined

Table 6.7: All possible combinations of attribute levels (1=low, 2=medium, 3=high), cohesion scores, balanced teams, and team averages (TAs) classified as low (L), medium (M), and high (H).

P1	P2	P3	Cohesion	Tot High	Tot Low	Bal?	TA	TA class
1	1	1	0.00	0	3	N	1.00	L
1	1	2	0.58	0	2	N	1.33	
1	2	2	0.58	0	1	N	1.67	
1	1	3	1.15	1	2	N	1.67	
2	2	2	0.00	0	0	Y	2.00	M
2	2	3	0.58	1	0	N	2.33	
1	2	3	1.00	1	1	Y	2.00	
3	3	1	1.15	2	1	N	2.33	
3	3	3	0.00	3	0	N	3.00	H
3	3	2	0.58	2	0	N	2.67	

above. For clarity, we display all possible high or low attribute distributions between four teams in Table 6.6. Unlike Odo [429], we classify these results into *evenly distributed* (when ED=0.43) and *unevenly distributed* (when ED>0.43).

Cohesion. Cohesion within a team is linked to the uniformity of attributes among team members. Our study uses Cohesion to address RQ4.3 and test H1.2 (see Table 6.4). According to Odo [429], high team cohesion correlates with a smaller standard deviation of attribute values within the team. To facilitate this calculation, we code high attribute values as 3, medium values as 2, and low values as 1, respectively (as shown in the second example of Table 6.5). to illustrate how Cohesion is computed, we provide an example of four teams with different Conscientiousness values:

- Team 1: 3,2,1 (1 high, 1 medium, 1 low)
- Team 2: 2,3,1 (1 medium, 1 high, 1 low)
- Team 3: 2,1,3 (1 medium, 1 low, 1 high)
- Team 4: 2,2,2 (3 mediums)

Cohesion is calculated as the standard deviation of the Cons values within each team. A lower standard deviation signifies better cohesion, indicating more uniform attribute levels within the team. The Cohesion for this set of teams is:

- Team 1 Cohesion (sd): 1
- Team 2 Cohesion (sd): 1
- Team 3 Cohesion (sd): 1
- Team 4 Cohesion (sd): 0

Therefore, in this example, the standard deviation for Cons is 1 for Teams 1,2, and 3, and 0 for Team 4, indicating a greater variation in attribute levels within the first three teams compared to the fourth. Within our study design, given the fixed set of attribute values (Section 6.5.2), there are four possible outcomes for Cohesion (Table 6.7), namely, 0, 0.58, 1.00, and 1.15. The lower the value, the higher the team Cohesion of the given attribute.

The Average Cohesion (AVGC) across all teams of a specific size (in this case, teams of three) is a weighted average of the Cohesion scores, where the weight for each score is the number of teams with that score. For example, from a pool of 100 teams, there are 30 teams with Cohesion=0, 30 with Cohesion=0.58, and 40 with Cohesion=1. The AVGC is calculated as:

$$AVGC = ((0 * 30) + (0.58 * 30) + (1 * 40))/100 = 0.57$$

The AVGC provides an overall measure of cohesion across all teams in the study. A lower AVGC indicates better overall cohesion among the teams, as it suggests that, on average, the attribute levels within each team are more similar.

Balance. Following Odo [429], we define balance as the equivalence between the quantities of high and low values for each attribute within a team. A team is seen as balanced for a given attribute if it accommodates an equal number of individuals with high and low scores. This metric can check for within-team attribute distribution and answer RQ4.2 (see Table 6.4). The count of medium (M) scores is not directly considered in determining balance as these values are considered neutral and do not tip the balance towards high or low attribute values.

In an example with four teams with three learners and three levels of Cons (high=H, medium=M, low=L):

- Team 1: H,M,L
- Team 2: M,L,H
- Team 3: L,H,M
- Team 4: M,M,M

Balance is calculated by comparing the count of high (H) and low (L) attribute scores in each team. If the count of high and low scores are the same, the team is considered balanced for the given attribute; if not, it is unbalanced.

- Team 1: H=1,L=1
- Team 2: H=1,L=1
- Team 3: H=1,L=1
- Team 4: H=0,L=0

All teams are balanced in the example since their high and low values are the same. Lastly, we compute the percentage of balanced teams, the number of balanced teams divided by the total multiplied by 100.

$$\text{Percentage of Balanced Teams} = \frac{\sum \text{Balanced Teams}}{\sum \text{Teams}} \times 100$$

Existing literature, such as the study by Curşeu et al. [119], posits that having average characteristics within a team can benefit effective team formation. Nonetheless, aiming for balance on one attribute could compromise the balance on another, given the fixed team size.

Team Average. In this case, we define a metric called Team Average (TA), the average value of a particular attribute and Cons for a given team. We use this metric to answer RQ4.2 (see Table 6.4). The calculation of TA can be explained as follows. Consider four teams with three learners each, with three possible levels for the Cons attribute (3=high, 2=medium, 1=low):

- Team 1: 3,2,1 (Average = 2.00)
- Team 2: 3,2,1 (Average = 2.00)
- Team 3: 3,2,1 (Average = 2.00)
- Team 4: 2,2,2 (Average = 2.00)

We then classify each team's average Cons value as high, medium, or low based on the lookup table in Table 6.7. Finally, we categorize each team's rounded average Cons value as high, medium, or low based on the lookup table in Table 6.7, where values are rounded to the whole number.

Finally, the percentage of TAs formed by the participants is calculated as follows.

$$\%\text{highConsTA} = \frac{\sum \text{highConsTA}}{\sum \text{teams}} \times 100$$

This measure gives us the overall distribution of teams based on the average attribute values, which is –in the case of the given example – Cons.

Attribute Distribution Thematic Frequency. Adapted from thematic analysis [67], the Attribute Distribution Thematic Frequency (ADTF) we define as the frequency of key terms related to attribute distribution in response to the survey question *Explain why you teamed the learners the way you did.* This metric is designed to gauge the prevalence of specific team formation strategies among respondents.

Firstly, we pre-processed the open-ended responses by removing stopwords and tokenizing each answer. This process reduced each reaction to a set of significant words or tokens. Then, we counted key terms related to attribute distribution: *balance, medium, equal,* and *average.*

The frequencies of these terms, denoted as f_{term}, were computed as follows:

$$f_{\text{term}} = \text{Count of the term term in the tokenized responses}$$

Table 6.8: Demographic information of the Main study participants (N=102) including Age, Employment Status, Gender, Student status, and Country of Residence.

Age	**N**	**%**
<18	0	0%
18-25	46	45%
26-33	36	35%
34-41	12	12%
42-49	4	4%
50-57	0	0%
58-65	1	1%
N/A	3	3%
Gender	**N**	**%**
Female	48	47%
Male	51	50%
N/A	3	3%
Student?	**N**	**%**
N/A	49	48%
Yes	22	22%
No	31	30%
Employment Status	**N**	**%**
N/A	47	46%
Full-Time	25	25%
Unemployed	11	11%
Part-Time	10	10%
Other	7	7%
Unpaid work	2	2%
Country of Residence	**N**	**%**
Portugal	25	25%
Italy	19	19%
Poland	10	10%
Spain	5	5%
Czech Republic	4	4%
Greece	4	4%
Germany	4	4%
UK	3	3%
Estonia	2	2%
Belgium	2	2%
Hungary	2	2%
Sweden	2	2%
Latvia	1	1%
Switzerland	1	1%
Finland	1	1%
Austria	1	1%
South Africa	10	10%
Mexico	1	1%
Israel	2	2%
N/A	3	3%

6

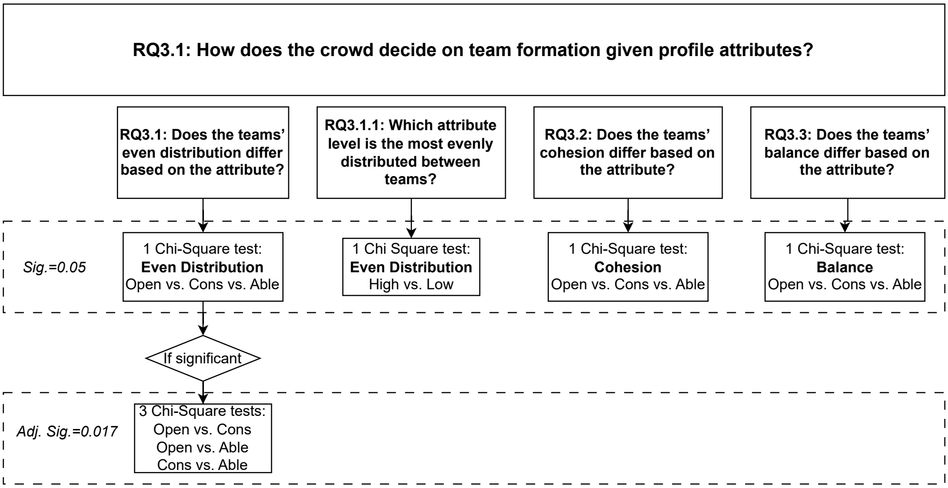


Figure 6.3: Overview of the Research Questions (RQ4, RQ4.1, RQ4.1.1, RQ4.2, RQ4.3) mapped to the statistical tests and significance levels.

Table 6.9: Even Distribution (ED) mean ED, standard deviation (SD), count (N), and percentage (%) of evenly- (ED=0.43) and unevenly-distributed (ED>0.43) teams for the three attributes Openness to Experience, Conscientiousness, and Ability.

Attr. Level	Avrg. ED	SD (ED)	N Even	% Even	N Uneven	% Uneven
Openness	0.55	0.21	150	74%	54	26%
Open **(High)**	0.52	0.20	84	82%	18	18%
Open **(Low)**	0.59	0.23	66	65%	36	35%
Cons	0.61	0.22	121	59%	83	41%
Cons **(High)**	0.56	0.22	72	71%	30	29%
Cons **(Low)**	0.65	0.23	49	48%	53	52%
Ability	0.59	0.22	129	63%	75	37%
Ability **(High)**	0.57	0.22	70	69%	32	31%
Ability **(Low)**	0.61	0.22	59	58%	43	42%
Total	0.58	0.22	400	65%	212	35%
Tot **(High)**	0.55	0.21	226	74%	80	26%
Tot **(Low)**	0.62	0.23	174	57%	132	43%

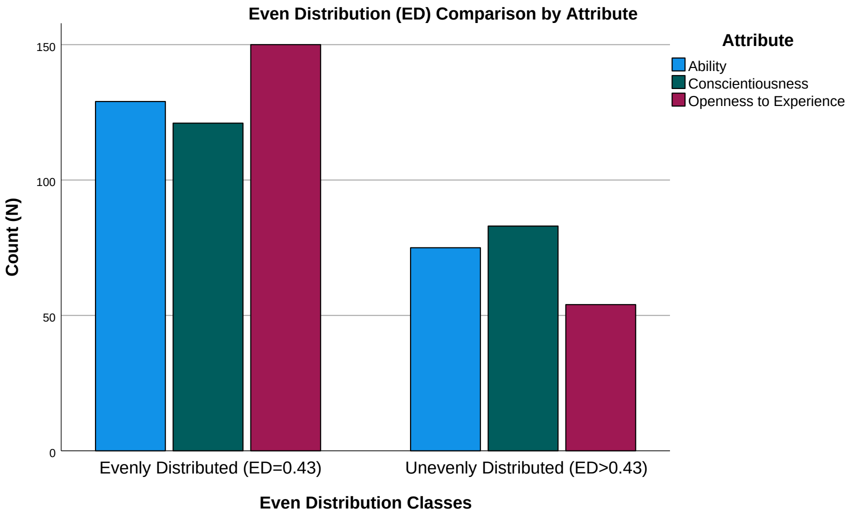


Figure 6.4: Even Distribution (ED) between attributes (Openness to Experience, Conscientiousness, and Ability). ED is classified as either evenly distributed (ED=0.43) or unevenly distributed (ED>0.43).

6.5.4. Participants

Similar to the pre-study, we recruited participants through Prolific. Out of an initial batch (*N*=120), most (*N*=102) completed the task correctly. The remaining eighteen contributions were discarded as they did not execute the task correctly (i.e., forming teams with the incorrect size or incomplete). Participant compensation complied with Prolific's recommended minimum wage (6.18/hour GBP) [469]. On average, participants spent 10 minutes on the task and thus received approximately 1 GBP each. More about the demographic characteristics of the participants can be seen in Table 6.8.

6.5.5. Results

Figure 6.3 gives an overview of the Research Questions and related statistical tests. This section presents the results of the descriptive statistics and the statistical tests starting from RQ4.1 and RQ4.1.1 (Sections 6.5.5 and 6.5.5), followed by RQ4.2 (Section 6.5.5) and RQ.3.3 (Section 6.5.5). We also present the collateral results from a post hoc analysis of the average attribute levels and the analysis of the participant's responses (Section 6.5.5).

RQ4.1: Does the teams' even distribution differ?

As discussed above, the Even Distribution (ED) metric quantifies how evenly attributes are spread across the four teams each participant forms. We classify ED values of 0.43 as Evenly distributed and higher values as Unevenly distributed. Table 6.9 shows the descriptive statistics for the Overall ED whereby the attributes low and high results are aggregated. Figure 6.4 displays the count of evenly and unevenly distributed teams by attribute.

More than half of the teams formed were evenly distributed on one or more traits (N=400, 65%), revealing a general preference for even distribution. When looking at the attribute type, Openness to Experience's mean ED is the smallest (mean ED=0.55, sd=0.21), indicating a better even distribution than Ability (mean ED=0.59, sd=0.22) and Conscientiousness (mean ED=0.61, sd=0.22).

We employed a Pearson chi-square test for significant differences in ED between attributes. The test result, $\chi^2(df) = 9.714, p = .008$, confirms a significant association between the even distribution and the attribute type. The Phi and Cramer's V were used as effect size measures. Results indicated a significant, small positive association $\phi = 0.126$, $p = .008$, and $V = 0.126$, $p = .008$.

To ascertain the differences between attributes, we ran a family of pair-wise Pearson chi-square tests. We corrected for Type I error by setting a new significance level (alpha) to 0.017 using a Bonferroni adjustment[4].

- *Openness to Experience vs. Conscientiousness - ED.* The test result, $\chi^2(df) = 9.242, p = .002$, establishes a significant association between the even distribution and the two attributes. The Phi and Cramer's V were used as effect size measures. Results indicated a significant, small negative association between Openness to Experience and Conscientiousness, $\phi = -0.151$, $p = .002$, and $V = 0.151$, $p = .002$. These results remain significant even after the Bonferroni correction.

- *Openness to Experience vs. Ability - ED.* The Pearson Chi-Square test result, $\chi^2(1) = 4.999, p = .025$, establishes a significant association between the even distribution and the two attributes. The Phi and Cramer's V were used as effect size measures. Results indicated a significant, small negative association between Openness to Experience and Ability, $\phi = -0.111$, $p = .025$, and $V = 0.111$, $p = .025$. However, these results do not remain significant after applying the Bonferroni correction.

- *Conscientiousness vs. Ability - ED.* The Pearson Chi-Square test result, $\chi^2(1) = .661, p = .416$, does not establish a significant association between the even distribution and the two attributes. The Phi and Cramer's V were used as effect size measures. Results indicated a minimal positive association between Conscientiousness and Ability, $\phi = 0.040$, $p = .416$, and $V = 0.040$, $p = .416$. However, these associations are not statistically significant as the p-values are far above the Bonferroni corrected significance level.

In summary, most participants prefer to evenly distribute attributes between teams, with **Openness to Experience being the most commonly evenly distributed trait**.

RQ4.1.1: Which attribute level is the most evenly distributed between teams?

Table 6.9 and Figure 6.5 display the ED results for high and low attribute levels. Overall, the total number of evenly distributed high attribute levels (N=226) exceeds that of the

[4]The adjusted significance derives from the division of the standard significance level (0.05) by the number of pair-wise tests (3).

6

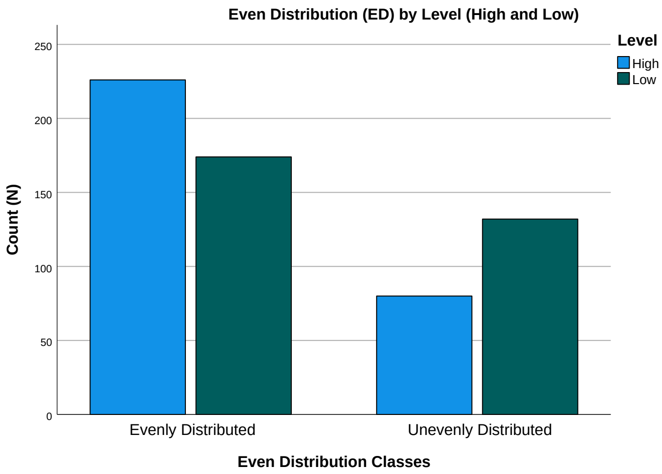


Figure 6.5: Even Distribution (ED) between attribute levels (High and Low). ED is classified as either evenly distributed (ED=0.43) or unevenly distributed (ED>0.43).

evenly distributed low attribute levels (N=174), showing that high levels tend to be more evenly distributed than low ones. Considering differences between attribute types, the high Openness to Experience mean ED (mean ED=0.52, sd=0.20) is better than that of high Conscientiousness (mean ED=0.56, sd=0.22) and high Ability (mean ED=0.57, sd=0.22).

6

A Pearson chi-square test explored the significant differences between high and low attributes. The test yielded a significant association, with $\chi^2(1) = 19.515$, $p < .001$. The effect size measures Phi, and Cramer's V was also calculated. The results indicated a positive association between the level of the attributes and the % of ED, with $\phi = 0.179$, $p < .001$, and $V = 0.179$, $p < .001$.

In summary, **high attribute levels tend to be more evenly distributed across teams than low attribute levels**. Among these, Openness to Experience is the most evenly distributed. Statistical tests reveal a significant association between high and low levels of these attributes, indicating that these two levels tend to go hand in hand regarding their distribution across teams. However, this is mainly determined by our study design and the distribution of traits in the learner pool.

RQ4.2: Does the team's cohesion differ based on the attribute?

Cohesion is calculated as the standard deviation of a team's high and low attribute values. Table 6.10 shows the descriptive statistics for this metric, including Cohesion average, standard deviation, and count of highly cohesive and poorly cohesive teams in the form of numbers and percentages. Highly cohesive teams have a Cohesion smaller than 1, while Poorly cohesive ones are greater or equal to 1. Figure 6.6 shows the count of highly and poorly cohesive teams by attribute. Overall, the results show that highly cohesive teams occur more often (N=720, 59%) than poorly cohesive ones (N=504, 41%). Diving deeper into the descriptive statistics, we note that Conscientiousness has

Table 6.10: Cohesion Average, standard deviation (SD), count (N), and percentage (%) of highly cohesive and poorly cohesive teams for the three attributes Openness to Experience, Conscientiousness, and Ability.

Attr.	Avg. Coh	SD	High Coh.	% High Coh.	Poor Coh.	% Poor Coh.
Openness	0.71	1.40	219	54%	189	46%
Cons	0.69	1.02	256	63%	152	37%
Ability	0.70	1.37	245	60%	163	40%
Total	0.70	1.26	720	59%	504	41%

6

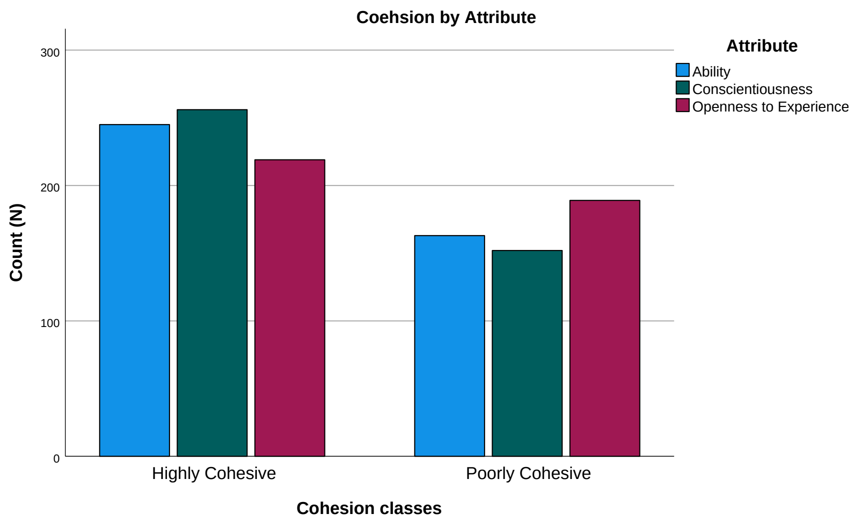


Figure 6.6: Cohesion between attributes (Openness to Experience, Conscientiousness, and Ability). Cohesion is classified as either highly cohesive (Cohesion<1) or poorly cohesive (Cohesion>=1).

Table 6.11: Distribution of Balanced and Unbalanced Attributes by Openness to Experience, Conscientiousness, and Ability.

Attribute	Balanced (N)	Balanced (%)	Unbalanced (N)	Unbalanced (%)
Openness	211	52%	197	48%
Conscientiousness	149	37%	259	63%
Ability	166	41%	242	59%
Total	526	43%	698	57%

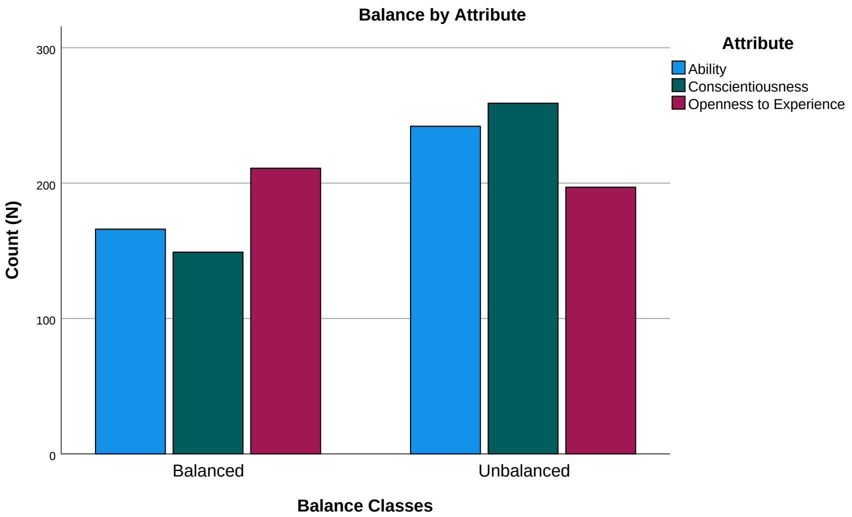


Figure 6.7: Balance between attributes (Openness to Experience, Conscientiousness, and Ability). Balanced teams are those with equal numbers of low and high attribute levels.

the most highly cohesive teams (N=256, 63%), followed by Ability (N=245, 60%) and Openness to Experience (N=219, 54%).

A Pearson chi-square test was performed to examine the significant differences between the levels of cohesion across the different attributes. The test yielded a statistically significant association, with $\chi^2(2) = 7.306$, $p = .026$. Symmetric measures further quantified this relationship, with Phi and Cramer's V providing a value of .077, thus corroborating the existence of a statistically significant association, albeit with a very small effect size. These results reveal that **high cohesion in team formation is more preferred than poor cohesion. We also note that Conscientiousness tends to be one of the most cohesively distributed traits (where learners of the same Conscientiousness are matched together)**.

Table 6.12: Count of attribute Team Averages (TAs) within teams classified as either Low (Low TA), Medium (Medium TA), or High (High TA).

Attribute	**Low TA**	**Medium TA**	**High TA**
Openness to Experience	16	292	100
Conscientiousness	23	254	131
Ability	19	269	120
Total	58	815	351

RQ4.3: Does the team's balance differ based on the chosen attribute (i.e., Openness to Experience, Conscientiousness, and Ability)?

The Balance metric indicates whether a team has the same high and low values of a given attribute. Table 6.11 and Figure 6.7 show the results for balanced and unbalanced teams by attribute. Unbalanced teams on at least one attribute (N=698, 57%) were slightly more common than balanced teams (N=526, 43%). Regarding the attribute of Openness to Experience, the teams were almost as unbalanced as balanced, with balanced teams (N=211, 52%) slightly outnumbering the unbalanced ones (N=197, 48%). In the case of Conscientiousness, we found more unbalanced teams (N=259, 63%) than balanced ones (N=149, 37%). Also, for Ability, there were slightly more unbalanced teams (N=242, 59%) than balanced ones (N=166, 41%). These results indicate that the specific attribute considered (Openness, Conscientiousness, Ability) influences the team balance.

6

The Pearson Chi-Square test result, $\chi^2(2) = 20.530$, $p < .001$, suggests a highly significant association among the three attributes: Openness, Conscientiousness, and Ability. Phi and Cramer's V, calculated as effect size measures, indicated a moderate positive association between attribute type and balance. The calculated values were $\phi = 0.130$, $p < .001$, and $V = 0.130$, $p < .001$. These results suggest a highly significant and small positive association among the three attributes: Openness, Conscientiousness, and Ability.

In summary, teams tend to be more unbalanced overall, but the degree of balance depends on the attribute in question. **Teams show nearly equal balance and unbalance when considering Openness, but are more often unbalanced when it comes to Conscientiousness and Ability**. Statistical tests reveal a significant and moderate positive association among these three attributes.

Post-hoc Analysis

Team Average. Our study relies on the Team Average (TA) metric to evaluate the formation of participant teams concerning average similarity across teams for each attribute type. Specifically, the TA is computed as the arithmetic mean of the attribute values within a team. In this experimental design, we could discern six distinct TA scores from ten possible combinations of low (1), medium (2), and high (3) levels. These were subsequently categorized as low, medium, or high (see Table 6.7). The analysis of our results reveals a trend toward forming teams with medium TAs, suggesting a balanced distribution of attributes that yielded an average between 2.00 and 2.33. The majority

Term	Frequency	Examples
Balance	35	*I tried to **balance** everyone so that the skills were similar.*
Medium	20	*My goal was to form teams with a balanced mix of high and low attribute levels, aiming for a **medium** range overall. This way, the teams would combine strengths and weaknesses, fostering collaboration and mutual support.*
Equal	16	*My objective was to maintain an **equilibrium** in all of the groups according to each participant's characteristics.*
Average	14	*I grouped the learners by similar attribute levels to ensure an **average** distribution.*

Table 6.13: Frequencies of Key Terms Related to Attribute Distribution from the participants' responses after the team formation task, along with examples.

of teams (N=292) reported medium TAs for Openness to Experience, as opposed to a smaller number of teams with low (N=16) or high (N=100) TAs. This pattern also holds for the attributes of Conscientiousness and Ability (see Table 6.12). Such findings suggest a preference or tendency towards distributing attributes among the teams, indicating that individuals might not necessarily aim for teams with extreme attributes but prefer a blend of qualities.

Attribute Distribution Thematic Frequency. To complement our analysis, we present a qualitative analysis of the responses to the survey question, *Explain why you teamed the learners the way you did.* by the Attribute Distribution Thematic Frequency (ADTF) approach. After preprocessing the open-ended responses by removing stopwords and tokenizing each answer, we identified significant words or tokens that provided insights into the strategies employed by the respondents. Table 6.13 shows the frequencies of the key terms after removing stopwords, namely balance (N=35), medium (N=20), equal (N=16), and average (N=14). To further illustrate the context in which these terms were used, we present some examples from the data collected:

- *I tried to **balance** everyone so that the skills were somehow similar.*
- *My goal was to form teams with a **balanced** mix of high and low attribute levels, aiming for a **medium** range overall. This way, the teams would combine strengths and weaknesses, fostering collaboration and mutual support.*
- *My objective was to maintain an **equilibrium** in all groups according to each participant's characteristics.*
- *I grouped the learners by similar attribute levels to ensure an **average** distribution.*

These examples reflect the respondents' intentions to balance attribute distribution when forming teams (for 85 words used to express the intent). The term may be closer to the medium average of the attributes (as calculated with the Team Average metric) than that of balancing high and low attribute levels (as indicated by the Balance metric). Overall, the ADTF analysis reveals that **the majority of respondents prioritize**

maintaining a balance (or average) in attribute distribution when forming teams. By considering the examples provided, we gain a deeper understanding of the strategies employed by the respondents in their team formation process.

6.6. Discussion

This chapter seeks to understand and model team formation through the eyes of the crowd, their strategies, and decisions when matching different individuals into teams. It extends previous work [429] using learners' personality traits and ability levels by engaging a wider pool of participants in the decision-making process, considering personality traits and adding additional analysis.

Since this study is strongly influenced by the work of Odo [429], we consider it essential to define the differences in study design that set our contribution apart.

Firstly, we note that our sample comprises of *N=102* crowd workers who are larger and (typically) more demographically diverse than Odo's [2021] University students sample (where *N=26* participants partook in the first study, and *N=14* in the second). Secondly, our research contrasts with that of Odo's [2021] in that it targets crowd workers in a remote setting via a dedicated web interface, extracting user-centred guidelines for future team formation systems. In contrast, their study focused primarily on collaborative learning in face-to-face settings. Thirdly, our study presents learners' profiles and attributes in a generalized format not tied to specific character descriptions, unlike the Persona-based approach used by Odo [429].

Furthermore, our initial investigation is dedicated to the relevance of all Big-5 personality traits, eventually focusing on Openness to Experience, Conscientiousness, and Ability based on feedback from crowd workers. In contrast, Odo's [2021] study used Conscientiousness, Agreeableness, and Ability in the first study and Extraversion, Emotional Stability, and Ability in the second, without regarding Openness to Experience. Lastly, Odo's [2021] study requires participants to distribute 12 learners into teams of various sizes (3,4, and 6). Our study design focuses only on forming four teams of 3 learners. The contributions stemming from the results of our study design are discussed in the following sections.

6.6.1. Preference for profiling attributes - Conscientiousness and Openness

Drawing on our pre-study, it becomes apparent that out of the Big-5 personality traits, the crowd perceives **Conscientiousness and Openness as the most influential attributes in forming teams of learners**. This finding was derived from a pool of crowd workers (N=52) who ranked these traits on a five-point Likert Scale, with the means indicating a stronger preference for Conscientiousness (mean=4.05, sd=0.82) and Openness to Experience (mean=4.09, sd=0.99). Interestingly, the other three traits - Agreeableness, Extraversion, and Neuroticism - were not as prevalent, as reflected by their lower means, namely 3.84 (sd=0.89), 3.30 (sd=0.94), and 2.61 (sd=1.25) respectively. From these findings, we steered to profile learners in the subsequent User As Wizard study using Openness to Experience and Conscientiousness. Additionally, we introduced Ability as the third attribute. The outcomes of this investigation coincide with previous research, underscoring the role of Conscientiousness [429] and Openness to Experience in team-based and cooperative learning environments [583] as well as our previous work on emergency response teamwork (Chapter 3). This opens the doors to new research avenues, exploring the impact of these personality traits on team dynamics and their efficacy in learning environments and online team formation.

6.6.2. Preference for evenly distributed attributes, especially Openness to Experience

In this study, we investigated how different attributes are distributed among teams formed by crowd workers. The Even Distribution (ED) metric was used to quantify the dispersion of attributes. Teams were categorized as either Evenly distributed (ED=0.43) or Unevenly distributed (ED>0.43). More than half of the teams formed were evenly distributed on one or more traits, indicating a **general preference for even distribution**.

When looking at individual attributes, Openness to Experience was the most evenly distributed attribute compared to Ability and Conscientiousness. Statistical tests also revealed a significant relationship between the even distribution and the attribute type.

These findings provide insights into team formation behaviour and attribute preference. Firstly, the preference for even distribution may indicate a **desire for fairness and balance in team composition**. Secondly, the fact that **Openness to Experience is the most evenly distributed attribute** could suggest that it's a valued trait in team settings, potentially due to its links with creativity and adaptability [608].

6.6.3. Preference for cohesion on Conscientiousness and variance on Openness to Experience

As part of our study, we investigated the cohesion of teams based on varying attributes. Cohesion was measured as the standard deviation of a team's high and low attribute values. Teams with a cohesion value of less than a given threshold (i.e., Cohesion<1) were classified as highly cohesive. Those with a value equal to or greater were considered poorly cohesive. The results showed that teams are typically more cohesive (N=720, 59%) than not (N=504, 41%). Among the attributes, the most highly cohesive teams were for Conscientiousness (N=256, 63%), followed by Ability (N=245, 60%) and Openness to Experience (N=219, 54%). A Pearson chi-square test confirmed a significant association between the cohesion level and the attribute type, albeit with a small effect size.

The findings suggest that when forming teams, there is a **general preference for high cohesion**, implying that the crowd prefers teams where attribute levels among members are similar. The fact that **Conscientiousness results in the highest degree of cohesion** could be due to its fundamental role in teamwork, as conscientious individuals tend to be organized, dependable, and diligent - traits that could potentially drive similar individuals to work together [429].

The existence of a statistically significant, albeit small, association between cohesion and attribute type hints at nuanced differences in how various attributes affect team cohesion. This further supports the idea that different traits have different implications for team dynamics. Future research could explore why specific attributes lend themselves to higher cohesion and the potential impacts of this cohesion on team performance. The fact that Openness to Experience has the lowest proportion of highly cohesive teams may also suggest that participants tried to form teams with at least one creative teammate (a facet commonly associated with highly Open to Experience individuals).

6.6.4. Preference for diversity in specific attributes

Another measure of team formation that this study considers is the balance within teams based on different attributes, i.e., whether a team has an equal number of high and low values for a given attribute. The results revealed that teams were more often unbalanced (N=698, 57%) than balanced (N=526, 43%). The degree of balance, however, depended on the specific attribute under consideration. For Openness to Experience, the teams were almost evenly distributed, with a slight edge for balanced teams (N=211, 52%) over the unbalanced ones (N=197, 48%). In contrast, Conscientiousness and Ability had more unbalanced teams, with 63% and 59%, respectively, compared to their balanced counterparts, 37% and 41%. Pearson's Chi-Square test showed a highly significant association between the attributes of Openness to Experience, Conscientiousness, and Ability, with a positive effect size (Phi and Cramer's V = 0.130).

The results indicate that **the attribute under consideration does impact the balance within teams**. The near-equal distribution of balanced and unbalanced teams in Openness to Experience might suggest that teams value a mix of different perspectives and ideas. In contrast, the higher number of unbalanced teams for Conscientiousness and Ability could reflect teams preferring to have a majority of individuals with high levels of these attributes, possibly for the efficient accomplishment of tasks.

6.6.5. Tendency towards moderation and balance in team formation

Aside from the main results, a post hoc analysis further evaluated the formation of teams using the Team Average (TA) metric. This metric calculates the arithmetic mean of the attribute values within a team. The study found a tendency towards forming teams with medium TAs, indicating a balanced distribution of attributes within teams. This was seen across the attributes of Openness to Experience, Conscientiousness, and Ability, with most teams falling in the medium TA category.

Additionally, a qualitative analysis using the Attribute Distribution Thematic Frequency (ADTF) approach was performed on survey responses asking why participants formed their teams the way they did. The study identified significant recurring words or tokens that shed light on the participants' team formation strategies. Four key terms emerged - balance, medium, equal, and average, revealing a common theme of aiming for balanced attribute distribution in teams. The preference for medium team averages suggests a tendency towards moderation and balance in team formation strategies.

Participants seem to prefer a blend of qualities rather than leaning towards teams with extreme attributes. This finding could be interpreted as recognizing the benefits of diversity and a balanced distribution of attributes within a team, allowing for more dynamic and multifaceted team performance. The qualitative ADTF analysis supports this finding. The recurring terms suggest that participants strive for balance or an average attribute distribution when forming teams. This indicates an understanding and value placed on the balance of different attributes within a team. These insights could have important implications for team formation strategies in various settings. They underline the importance of considering a balanced mix of attributes when forming teams rather than focusing on extreme attributes. This could potentially lead to more effective team dynamics and improved team performance. Furthermore, these results could be leveraged in developing online team-building training programs and explainable team formation algorithms. Such implementations could aim to enhance participants' understanding of the importance of a balanced attribute distribution within teams, which could, in turn, facilitate more effective team formation and collaboration.

6.7. Limitations

While our study provides valuable insights into how individuals approach team formation tasks given different learner profiles, several limitations should be considered when interpreting the results.

- **Artificial Context:** The first limitation pertains to the artificiality of the study context. Our study relies on fictitious learner profiles, which may help to control the study design but not accurately reflect the complexity and variation of real-world individual profiles. Moreover, this design could limit the participants' ability to empathize or deeply understand the profiles, impacting their team formation strategies.

- **Limited Attributes:** Our learner profiles were characterized by only three attributes: Conscientiousness, Openness, and Ability. Although significant in team formation, these attributes do not encompass all possible attributes that might influence team dynamics and performance, such as experience, other personality traits, skills, and preferences. Thus, the results may not fully reflect the complex factors in real-world team formation scenarios.

- **Attribute Presentation:** How we presented attributes might have influenced the decision-making process. All attributes were displayed with the same visual prominence, which might have unintentionally suggested that they all hold equal importance. This may not align with participants' perceptions or how these attributes affect team performance.

- **Sample Bias:** Our participant pool was sourced from Prolific, which may not represent the general or crowd-working population. Our findings may not generalize to other subjects with different cultural, educational, or professional backgrounds.

- **Limited Validation of Qualitative Responses:** Our analysis of the qualitative responses to the open-ended question relies heavily on the honesty and accuracy of the participants' self-reported strategies. We can only take these responses at face value without further validation measures such as follow-up interviews or observational data.

- **Team Size:** The team size in this study was fixed at three. The dynamics and considerations for forming larger or smaller teams may differ significantly.
- **Statistical analysis and correction methods:** Despite the insights offered by this study, the statistical approach used to analyze the results imposes certain limitations. Firstly, our use of Pearson's chi-square test to identify differences assumes the data to be categorical and the observations to be independent, which may not wholly hold in our case. This assumption could potentially influence the robustness of our findings, requiring caution in interpreting the results. Secondly, the application of the Bonferroni correction, while effectively controlling the Type I error rate across multiple comparisons, may be overly conservative, leading to a decrease in statistical power and potentially causing us to overlook significant associations (Type II error). This means that some critical differences between attributes may not have been detected.

Despite these limitations, our study offers a valuable starting point for understanding individual strategies in team formation tasks. Future research could address these limitations by employing more comprehensive and realistic learner profiles, considering a broader range of attributes, and employing a more diverse participant sample. Overall, while the choice of metrics provides a sound starting point for investigating team formation in crowd-based systems, they represent a simplified view of the complex dynamics of real-world teams. Additional qualitative or mixed methods research could be valuable in complementing these metrics and in better understanding the nuances of team composition and dynamics.

6.8. Conclusion

This chapter was dedicated to understanding the crowd's perception of team formation and their strategies for grouping individuals based on specific profile attributes. Drawing from a User-Centered approach, we engaged 102 crowd workers in a team formation experiment simulating an online education scenario. With learners' profiles defined by Conscientiousness, Openness to experience, and cognitive ability levels, the crowd was tasked to form four teams of three learners. Through this work, we observed the preference for conscientiousness and openness to experience when forming teams, which aligned with previous studies. Moreover, our study suggests that teams formed by crowd workers generally prefer an even distribution of attributes, particularly for Openness to Experience. Interestingly, we observed a tendency towards forming teams with a balanced distribution of attributes, indicating an inclination towards moderation in team formation. The following chapter discusses the thesis' findings and provides recommendations for future work.

7

Discussions and Conclusions

This thesis explored and defined some of the most critical factors in collaborative crowdsourcing systems: crowd workers' preferences, behaviour, and decision-making in various collaborative settings. **The result is a collection of findings from a diverse pool of user studies that can guide the improvement of collaborative crowdsourcing systems**. The field of crowdsourcing has seen significant evolution. However, the design of crowdsourcing systems frequently overlooks the unique differences and needs of the individuals who utilize them (e.g., [224, 72]). The consequences are often dire as crowd workers' preferences, opinions, and unique capabilities remain capped, affecting their sense of ownership, social recognition, and affiliation with their work [357]. This thesis proposes addressing the following *risks* associated with online crowdsourcing collaborative environments that fail to capture crowd workers' requirements:

1. *Neglecting crowd workers' perspectives.* By not considering the opinions and preferences of crowd workers, there is a missed opportunity to improve system trustworthiness and user engagement.

2. *Overlooking the validation of system assumptions.* Failing to verify the in-built assumptions of the system (e.g., how to profile crowd workers, how to recommend or match them to others, how to decide which attributes are important for the specific task, etc.) can lead to subpar outcomes, potential discrimination, and dissatisfaction.

3. *Disregarding the impact of individual and team characteristics and dynamics.* Ignoring the influence of personal and team attributes on collaboration and satisfaction can result in overlooked opportunities for process optimization and more effective team formation.

In this concluding chapter, we revisit the scope of this thesis and highlight the repercussions of our findings. Finally, we conclude with a list of the main limitations and directions for future work.

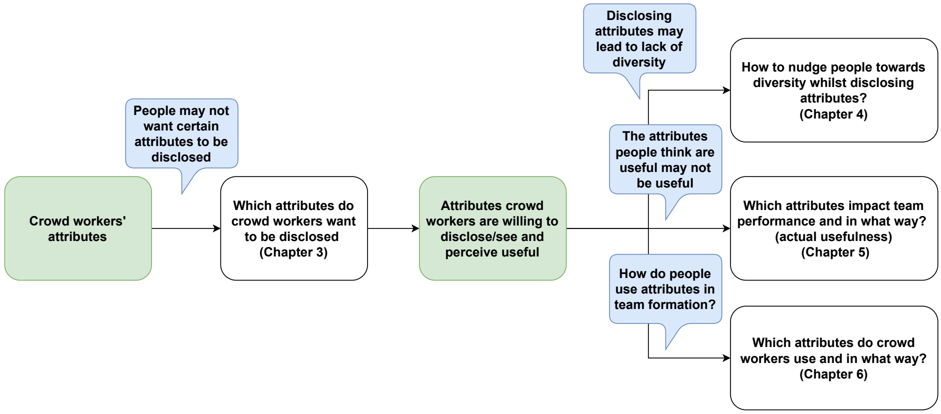


Figure 7.1: Overview of the thesis structure.

7.1. Towards the Future of Crowdsourcing User-Centered Teams

Figure 7.1 illustrates the thesis structure. The main Research Question defines the whole research framework:

RQ: What are the critical factors for developing user-centred collaborative crowdsourcing systems that promote engagement, efficiency, and diversity in online crowd teamwork?

While the literature presents ample solutions for user-centred crowdsourcing systems (e.g., taxonomies in [188, 493, 130]), limited research specializes in user-centred collaborative crowdsourcing, where participants actively work together. Frameworks like De Vreede et al. [130] explore user engagement (level and quality of effort) in open collaboration, highlighting factors like personal interest, clear goals, and intrinsic motivation as critical drivers. This thesis builds on this foundation by examining how crowd workers' preferences, traits, team dynamics, and decision-making shape our understanding of their needs and viewpoints in collaborative settings. This thesis dives into four main research areas:

1. **Chapter 3:** The preferences of crowd workers for surface- and deep-level attributes in team assembly.
2. **Chapter 4:** The influence of digital nudging interventions on selecting diverse teammates for collaborative outsourced projects.
3. **Chapter 5:** The influence of personality traits and communication styles of individuals and teams on collaborative tasks, especially under stress.
4. **Chapter 6:** Strategies employed by the crowd to allocate individuals of varying deep-level attributes such as personality traits and ability levels into teams.

These studies collectively inform the design for user-centred collaborative crowdsourcing systems, emphasizing the importance of understanding the users to provide more engaging, efficient, and diverse online crowd teamwork. Therefore, this thesis proposes a list of critical factors and methodologies for developing user-centred collaborative crowdsourcing systems, focusing on user engagement in system design choices, team efficiency based on compatibility between individuals and tasks, increased diversity in online team assembly, and elicitation of users' decision-making strategies in team formation. The identified *critical factors* are:

1. **Crowd workers' preferences** for certain surface- and deep-level attributes in team formation (Section 7.2).
2. **Digital manipulations** such as nudging interventions to drive crowd workers towards team diversity (Section 7.3).
3. **Individual's and team's personality traits and communication styles** for effective team collaboration, especially in high-pressure situations (Section 7.4).
4. **Decision-making of crowd workers** when considering different kinds of profiling attributes during team formation (Section 7.5).

These findings contribute to creating more dynamic, inclusive, and user-centric crowdsourcing systems. In the following sections, we discuss the research contributions yielded from each chapter and embed them into the broader landscape of the recent literature on collaborative crowdsourcing.

7.2. Crowd workers prefer specific attributes

An initial step in enhancing crowdsourcing systems for collaboration involves *determining how crowd workers' profiling characteristics should be made visible within online systems.* In Chapter 3, we investigated crowd workers' preferences for sharing information in the context of online crowdsourcing team formation systems, focusing on the types of profiling attributes they are comfortable seeing and showing. The study distinguishes surface-level and deep-level traits [304]. Surface-level traits include easily observable characteristics such as age and gender, while deep-level traits cover less visible attributes such as beliefs and attitudes. The following research questions guided the study design and the analysis of the results:

RQ1: Which personal and professional profile attributes do crowd users prefer to see and show on crowdsourced team formation systems?

RQ1.1: *About themselves*, which personal and professional profile attributes do crowd workers prefer to display on crowdsourced team formation systems?

(a) Which *types* of attributes (surface-, deep-level) are crowd workers *willing to display* about themselves?

(b) Which *types* of attributes (surface-, deep-level) do crowd workers *find useful* to display about themselves?

(c) Are crowd workers *more willing to display* surface- or deep-level attributes about themselves?

(d) Do crowd workers find it *more useful to display* surface- or deep-level attributes about themselves?

(e) Which *individual* attributes are crowd workers *willing to display* about themselves?

(f) Which *individual* attributes do crowd workers *find useful to display* about themselves?

RQ1.2: *About others*, which personal and professional profile attributes do crowd workers prefer to see on crowdsourced team formation systems?

(a) Which *types* of attributes (surface-, deep-level) are crowd workers *willing to see* about others?

(b) Which *types* of attributes (surface-, deep-level) do crowd workers *find useful* to see about others?

(c) Are crowd workers *more willing to see* surface- or deep-level attributes about others?

(d) Do crowd workers find it *more useful to see* surface- or deep-level attributes about others?

(e) Which *individual* attributes are crowd workers *willing to see* about others?

(f) Which *individual* attributes do crowd workers *find useful to see* about others?

In a survey-based study with 117 crowd workers from Prolific [469], we found a significant preference for sharing surface-level characteristics such as age, education, and social media information (i.e., availability, profile photo, and popularity). These findings align with the crowd workers' perceived usefulness where practical, less sensitive information is desired. Conversely, mental states, beliefs, and political affiliations are significantly worse, as they are likely viewed as too personal or sensitive. Another important finding is that crowd workers are willing to see and perceive useful specific deep-level traits such as personality, values, opinions, and topical interests. Finally, we observed no notable difference in how much individuals were willing to share their information and how much they were willing to see the same information on other crowd workers' profiles within the team formation systems. A similar result applies to the perceived usefulness metric. This suggests a consistent standard in the perceived usefulness and appropriateness of sharing personal profiling attributes among participants, regardless of whether the information pertains to themselves or others in the same environment.

These findings indicate that carefully pairing surface-level attributes with more specific deep-level ones aligns with crowd workers' preferences and may facilitate their decision-making and judgement when seeking teammates online. Additionally, the results suggest that displaying sensitive information such as ethnicity, religion, and depression on crowdsourcing platforms goes against what the crowd wants, a sign that the average crowd workers may prefer avoiding disclosing information that can be used against them and subject to anti-social behaviour and discrimination in the workplace [268]. These results align with earlier research examining how posting culturally significant personal information online can fuel discriminatory behaviour [2, 6]. Regarding the crowd workers' adversity to disclose online certain deep-level traits such as ethnicity, religion, and depression, we suspect this may be determined by the nature of the platform (i.e., online community and digital labour) and level of trust in the system.

While sharing basic, surface-level attributes is generally considered the safest approach to minimize negative behaviours, it is essential to recognize that virtually any type of profiling information—including seemingly innocuous details like gender, age, and profile photos—carries the potential for misuse. Unfortunately, foreseeing the negative consequences of revealing what appear to be harmless attributes poses a significant challenge. The research by Abramova [2] highlights this issue on shared economies platforms. In their use case, drivers and co-travellers with male names of Middle Eastern descent show a comparative disadvantage as they receive –on average– fewer offers than those with typically Western names. This, and similar studies on the adverse effects of disclosing information online [6, 335] exemplify how revealing specific attributes can facilitate collaboration and trust in online settings whilst also triggering harmful and discriminatory behaviour. In summary, our literature research indicates that disclosing crowd workers' most acceptable and relevant attributes and skills for the task is a complex issue, as it can result in positive and negative consequences. The following paragraphs discuss the pros and cons of sharing information online for crowdsourcing collaborative work.

Increasing social transparency benefits the individuals and the team. While disclosing information online presents its risks [2, 6, 335], research has shown that increasing social transparency by revealing personal information benefits online teams and even makes a (positive) difference in their work outcomes. In collaborative crowd work, sharing basic personal details among team members, such as names and nationalities, can significantly enhance teamwork [251]. The study by Huang and Fu [251] suggests that when team rewards depend on the team's overall performance, displaying information such as name and nationality reduces the likelihood of members not pulling their weight (i.e., social loafing) as everyone becomes more accountable to the team. Additionally, in competitive settings where rewards are based on outperforming others, showing the same information can motivate workers to excel, leveraging the natural drive to perform better in the presence of peers (i.e., social facilitation). From our results, displaying gender, age, topical interests, education, opinions, values, personality, and other social-media-related information (i.e., availability, profile photo, rating, popularity) is – at least from the perspective of crowd workers – acceptable and desirable since it may be used to enhance the search for the most compatible and adequate teammate(s).

Deciding which attributes to share determines whether crowd workers may be excluded or alienated. Sharing information about crowd workers online, such as language proficiency, achievements and job completion rates, aims to benefit them. However, for this information sharing to be truly effective, it must be personalized and contextualized to meet individual preferences and needs. Merely displaying "basic" attributes and task-related achievements might not give crowd workers a strong sense of ownership and identity, which are crucial for boosting their engagement in crowdsourcing platforms. The research by Munoz et al. [408] highlights the challenges online freelancers face, who feel that digital labour markets overly focus on client satisfaction and platform ratings rather than the workers' individual stories and experiences.

For many online workers, managing their reputation through client feedback and ratings becomes central to how they present themselves. Their study suggests that a freelancer's identity on these platforms is often reduced to a standardized profile dominated by skills, ratings, and metrics, heavily influenced by the platform's algorithms. The standardization, along with strict platform policies and evolving designs, controls and limits workers' identities, leading to what Munoz et al. [408] describes as *"a form of indentured servitude"*. This critique calls for a reassessment of how digital platforms manage and present worker information, advocating for practices that recognize and respect the individuality and autonomy of online workers.

Based on our research and the review of related literature, we recommend conducting studies to assess both the immediate and extended impacts of revealing profiling attributes within crowd team formation platforms. Additionally, we emphasize the crucial role that the specific context and nature of the task play in how these attributes are used—or potentially misused.

7.3. (Certain) Digital Interventions promote diversity

Following the conclusions from Section 7.2, we set off to *evaluate ways for systems to moderate the adverse effects of social transparency*. By social transparency, we intend the system and people's openness to share information. In our case, this information concerns crowd workers, their backgrounds, knowledge, and preferences, which can be made visible on team formation online platforms. To investigate *how* to moderate the adverse effect of social transparency in online crowd work settings, we used a between-subjects experimental study design reported in Chapter 4. Revealing crowd workers' characteristics, such as gender and race, could impact how others, such as collaborators and employers, perceive them. A race and gender-bias study on two prominent online freelance marketplaces (i.e., types of crowdsourcing labour) - TaskRabbit and Fiverr [222] demonstrated that displaying gender and race can significantly limit the opportunities to obtain work. In particular, the study shows that despite serving different markets (TaskRabbit is tailored for physical tasks and Fiverr for virtual ones), both platforms exhibit a concerning trend of biased social feedback against women and people of colour[1].

[1] On TaskRabbit, women, particularly White women, receive 10% fewer reviews than their male counterparts with similar work experience, while Black men face notably lower feedback scores. On Fiverr, Black men not only get 32% fewer reviews compared to other men but also score lower in ratings, a trend worse only for those without a profile image. Additionally, reviews for Black women contain fewer positive adjectives, with Black workers' reviews overall having more negative language. In contrast, Asian men on Fiverr are rated significantly higher than others.

Research in behavioural economics has demonstrated that an effective strategy to influence people's behaviour towards desired outcomes (e.g., choosing more diverse teammates) is not by limiting their choices but by steering them towards the most beneficial options [563]. This approach is broadly known under the umbrella term *nudging* interventions [562]. Our study proposes to extend the literature on nudging interventions – mainly digital nudging (i.e., design features which lead or encourage users to follow the designer's preferred paths in the user's decision-making [552])– to counteract the adversarial effects of showcasing various information about crowd collaborators online. The following research questions guided our study:

RQ2: What is the impact of digital nudging techniques on promoting diversity in self-assembled crowd project teams?

(a) *(How) does Priming affect the diversity of the members that crowd users select for their team?*

(b) *(How) does displaying Diversity Information (DI) affect the diversity of the members that crowd users select for their teams?*

(c) *(How) does the combination of Priming and Diversity Information (DI) (Priming + DI) affect the diversity of team members that crowd users select for their teams?*

In answering **RQ2** and promoting diversity in self-assembled crowd project teams, we tested the effects of two techniques and their combinations. The first technique displayed explicit personalized Diversity Information in the form of the current team Diversity Score (an aggregate measure in the form of a progress bar) and diversity recommendations (in the form of a colourful highlight on those profiles whose characteristics are the most different from the user's). The second technique used diversity priming in the form of counter-stereotypes and All-Inclusive Multiculturalism, manipulating the presentation of the outsourcing company leader and the recruitment slogan to emphasize the company's inclusive policy. For the evaluation of these nudging interventions, we asked 120 crowd participants working on a crowdsourced innovation project scenario recruited through Prolific and Amazon Mechanical Turk [469, 71]

Our results indicate that subtle and less direct nudging interventions, such as priming, can adversely affect diversity policies and may even deter users from picking teammates from different regions. We also observed that displaying Diversity Information in simple visual cues (e.g., progress bars and colourful tags) can nudge crowd workers into choosing more diverse teammates. Other factors we also found to predict selection behaviour were the participants' region of origin, gender, teammates' functional backgrounds, and their order of appearance. The following paragraphs discuss the pros and cons of relying on digital nudging interventions (and the different kinds) to foster inclusion and diversity in online crowdsourcing collaborative spaces.

Explicit aggregated scores drive diversity in team formation. Enhancing diversity through digital interventions requires more than subtle nudges; explicit, visual, and aggregate measures prove far more effective. Our study shows that such interventions significantly promote diversity when applied to crowd worker team selections. We introduced a Diversity Information condition, utilizing a progress bar visually representing team diversity through an aggregated score derived from the Blau Index, assessing the mix of chosen profile attributes. This explicit, aggregated visualization helps offset biases conditioning user behaviour during teammate selection and collaboration, steering towards more diverse and inclusive team compositions.

While displaying diversity through such an aggregated value presents significant limitations (as discussed in Chapter 4), it may also compensate for certain biases known to condition users when looking for teammates and collaborating with others of dissimilar characteristics. In their study on the effect of displaying gender and racial profiling information on a collaborative crowdsourcing task, Baten et al. [42] found that disclosing specific demographic cues such as age and gender of the crowd workers can lead to an increase in homophily in team formation and a decrease in diversity and novelty of creative outputs. When such demographic information was available, there was a significant increase in the formation and persistence of same-gender connections, but not same-race connections. Furthermore, inter-ego semantic similarity—how similar the ideas generated by different individuals were—increased significantly when teammates' demographic cues were known. These undesired effects can be mitigated when the diversity of the collaborators is displayed as an aggregated measure - for instance, through a progress bar indicating on a scale of 1 to 100 how much diversity there is in the team.

Priming toward diversity can backfire. The work by Bolton et al. [56] defines two possible ways to nudge people into changing their behaviour. One method is through observability, which capitalizes on the visibility of an individual's actions to others. Another technique is to deploy social and economic incentives without the element of being watched. *Observability nudges* are predicated on the idea that people will alter their behaviour in socially desirable ways when they know others are monitoring their actions due to concerns about reputation or social approval [56]. In contrast, nudges that use *social and economic incentives* aim to motivate behaviour change through direct benefits or penalties or by appealing to intrinsic social preferences, such as fairness or altruism, without necessarily making the individual's actions observable to others [56]. The priming technique used in our study strongly resembles the latter approach. As shown by Bolton et al. [56], efforts to nudge people towards more inclusive behaviours using social and economic incentives—like those we experimented with in our priming condition (involving All Inclusive Multiculturalism and counter-stereotypes) —might not always work as intended. Instead of fostering genuine inclusivity, these nudges can sometimes lead to token gestures or shallow compliance. Additionally, there is a risk that people might prefer to work with others who are like themselves in an attempt to balance what they see as typical behaviours within their company. This "homophily", or the tendency to associate with similar others, could inadvertently sabotage diversity by justifying choices that exclude those who are different, under the mistaken belief that it compensates for biases against certain groups. To promote diversity and moderate nudging interventions' adverse effects, Bolton et al. [56] suggests deepening the embedding of desirable behaviours into the organization's culture. This could include setting up mentorship opportunities, conducting diversity training that engages employees, and sharing diversity-related achievements and goals. Platforms like OpenIDEO and Kaggle could incorporate these and similar initiatives to support their communities.

Profiling attributes strongly influence crowd workers when choosing teammates. Profiling attributes significantly impact how crowd workers select their teammates in collaborative online environments. Chapter 3 establishes that crowd workers prefer to have specific profiling attributes visible online. They are comfortable with most surface-level characteristics like age or gender and certain deep-level attributes such as skills or education. However, attributes potentially linked to discrimination or social bias are less favoured. Chapter 4 delves deeper, examining the influence of specific profiling attributes on selecting teammates for outsourced tasks. It was found that participants often chose teammates of the same gender, demonstrating a tendency towards gender homophily. This behaviour aligns with previous research findings, suggesting that individuals naturally associate with others with similar characteristics [42]. Moreover, the functional background of potential teammates—defined by their skills, expertise, and professional experience relevant to the task at hand—played a critical role in the selection process. Profiles highlighting expertise in areas directly related to the project, such as sales and marketing for creating a project to create a coffee slogan, were notably more likely to be chosen. This suggests that while initiatives to promote diversity can sway decision-making, the overarching consideration for participants remains the task's requirements. From the crowd workers' perspective, effective team composition prioritises aligning team members' functional backgrounds (e.g., work experience and skills) with the project's goals. Additionally, the study observed that crowd workers' responses to efforts to increase team diversity, such as displaying Diversity Information, varied by their regional background. European participants, for example, were more receptive to such diversity nudging and showed a higher propensity to adjust their behaviour towards selecting more diverse teammates. A more granular analysis is necessary to discern if European receptiveness to diversity nudging masks underlying disparities in attitudes and behaviours among European countries as suggested by similar studies [464].

Design biases impact crowd worker's decisions. Some of the results, particularly the tendency for people to choose teammates from the top of the list, can be linked to established cognitive biases in decision-making processes, namely priming, anchoring, and availability biases [631]. These biases explain how individuals perceive, process, and prioritize information in decision-making contexts, such as selecting teammates in a digital platform [456]. Our results align with the possible presence of *priming effects* [426], where users' familiarity with traditional list layouts might predispose them to focus on and select options presented at the top. This bias is reinforced through repeated interactions with similar interface designs across various platforms, including search engines like Google, where top results gain disproportionate views [426].

The *anchoring effect*, which biases individuals towards initial pieces of information they encounter, can also explain part of our findings. Just as the first result on a Search Engine Results Page can disproportionately influence users' perceptions of subsequent information [426], the initial options in the teammate list likely served as anchors, skewing preferences towards those presented earlier and potentially affecting the diversity of team selection. Lastly, the *availability heuristic*, which suggests that people rely on immediately recalled information, might explain users' preference for top-listed options. Given the cognitive effort required to evaluate each potential teammate thoroughly, users may default to selecting readily available or recallable options – in this case, those at the beginning of the list. Considering these and several other cognitive biases, user interfaces designed for diversity and inclusion must include interventions to moderate their undesirable consequences, such as randomising listing, chunking and categorising the options, and user education on bias.

7.4. Undisclosed characteristics matter in teamwork under pressure

In several instances, crowd workers collaborating with others are unaware of each other's characteristics. While we have explicitly handled profiling attributes in previous chapters, in this study, we assess a use case whereby crowd workers are entirely unaware of their teammates' characteristics. In Chapter 5, we investigated *the impact of covert (i.e., hidden) factors such as personality traits and communication styles on crowdsourcing teamwork under stressful conditions* such as when teams need to solve urgent problems given limited time and information. The following research questions guided the study:

RQ3: How do personality, communication patterns, and other user characteristics affect online ad hoc teams under pressure in emergency response situations?

i) *What personality characteristics render high-stake online teams successful?*

i) *Which skills, abilities, or socio-cultural elements must be considered when forming these teams?*

i) *Are there any particular communication patterns that can serve as early signals of effective teamwork under stress?*

In answering **RQ3**, we explored the dynamics between 120 crowd participants in 60 virtual dyads and their collaboration outcomes during 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[2] more likely to defuse the bomb on time. An analysis of communication patterns suggests that winners used action and response statements more. The team role was linked to the individual's preference for specific communication patterns and related to their perception of 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 that gender may have some impact

[2] Minimum levels of Openness means that the team member with the lowest level of Openness to Experience within each team, on average, exhibited higher Openness compared to the lowest levels observed in other teams.

on performance. In the following paragraphs, we discuss the repercussions of these findings.

The nature of the task and the personality traits of the crowd teams go hand-in-hand. As seen from several other studies discussed in Chapter 3, our results confirm that individual and team characteristics play a significant role in teamwork outcome and perception under a specific type of task- namely, a stressful real-time cooperative task. In this part of the thesis discussions, we want to emphasise the nature of the task. While our results show that Openness to Experience matters in teamwork effectiveness, this finding is constrained by the unique nature of the task – completing a bomb-defusion cooperative activity through a novel interface. Considering the significant impact that the task has on the chances of crowd teams to succeed, we must refer back to the valence of designing systems that adhere to taxonomies on team types and task types (e.g., [247]) to optimize the assembly of efficient and effective teams.

Given the variable nature of tasks and team dynamics in open collaboration and crowdsourcing settings, designing adaptive systems that can flexibly accommodate different tasks and teams becomes critical – and likely highly challenging. These systems should be capable of assessing task demands in real time and adjusting team composition and strategies accordingly to optimize performance. The granularity of the approach to differentiating between task types and team types will depend on whether 1. crowd workers are willing to share relevant information (Chapter 3), 2. the information about the crowd is relevant (or, in some cases, essential) to the task (Chapter 4), 3. the level of ease of monitoring and gathering information about the crowd and its dynamics [241].

Clear communication styles and roles can make or break crowd teams. Communication is essential in teamwork [295] and can benefit individuals in various circumstances [499]. Previous research has shown that crowd workers seek communication and peer feedback – even for individual microtasks. It also shows that communication between workers (through blog posts and forums) significantly improves the outcome of the task and the requester's utility [557]. In high-pressure scenarios, like the virtual bomb defusal task studied, the ability of team members to communicate clearly and respond to each other's cues becomes critical. Action statements (direct instructions or information crucial for immediate next steps) and response statements (replies or confirmations to action statements) indicate a team's ability to engage in goal-oriented communication. This focused communication pattern helps ensure that all team members are aligned in understanding the task at hand and the actions required to complete it successfully. It minimizes misunderstandings and ensures that responses to challenges are swift and coordinated.

We suggest combining our findings on the relevance of communication styles in crowd cooperative emergency tasks with taxonomies of teamwork and communication similar to Serçe et al. [520]. Their work combines communication styles (influenced by the task and cultural and linguistic differences), modes (e.g., synchronous like chats or asynchronous like forums), task complexity (e.g., micro- vs. macro-tasks), and leadership (e.g., teams with experts and leaders tend to have more structured and goal-oriented communication) in globally distributed teams. Our results demonstrate that ad-hoc crowd teams can establish structured and goal-oriented communication given the right *communication channel* (synchronous mode), clear *role assignment* (unambiguous and interdependent), and *focused and coordinated task* objectives (shared goals and accountabilities). Moreover, our results suggest that the individual preference for specific communication patterns might be linked to the team roles, affecting the team's overall performance and each member's perception of the collaboration quality. This implies that in team settings, especially under stress, the compatibility of communication styles between crowd workers and their roles and between different roles in the team can significantly impact the results.

How crowd workers perceive collaboration depends on their personalities. As seen from our study, certain personality traits not only impact the effectiveness of the collaboration but also the individuals' perceptions of how the collaboration went. Crowd workers with high agreeableness coped better with losing, implying that agreeable individuals have a more positive perception of their team's collaboration quality, even in less favourable outcomes. In our study, crowd workers had limited communication time during the task. They were severely restricted by how much they could share regarding feelings, opinions, and other strictly non-task-related topics. While it was not possible to determine to what extent the individuals' perception of the collaboration affected that of the team and vice-versa, similar studies show that the individuals' attributes, such as Agreeableness and Conscientiousness, are moderated by team-level confidence perceptions (i.e., collective efficacy, or, the team's collective confidence in its ability to complete tasks, overcome challenges, and achieve its goals), which combined affect the collective efficacy [560].

Our study also shows that teams that were heterogeneous regarding Conscientiousness (i.e., teammates who had different levels of Conscientiousness) felt better about their collaboration quality. This suggests that diversity in certain personality traits within a team can enhance the team's overall perception of working together. This finding contrasts that of other studies with different types of teams and channels of communication since different Conscientiousness levels typically can cause a lowering in satisfaction. For example, the work by Gevers and AG Peeters [193] shows that, within student teams, individual-level differences in Conscientiousness can directly decrease team members' satisfaction with the team; however, these differences do not necessarily impact their satisfaction with the team's performance. On the other hand, team-level dissimilarity in Conscientiousness can have a negative indirect effect on both satisfaction types by complicating early agreement on task timelines, which, in turn, disrupts coordinated action in later stages. The comparison of the findings suggests that the individual's personalities and the context (including the task, the communication mode, and roles) are moderated by the levels of social contagion (i.e., how much people of the same team influence each other by sharing their experiences and opinions), which in turn affects individual and team-level perceptions and efficacy.

7.5. Prioritizing personality traits in team formation

Up until now, we have witnessed how crowd workers' attributes affect – one way or another – collaborative endeavours. Our final research aimed to observe *the crowd's approach to team formation as a task*. We wanted to see how crowd workers handle different profiles and attributes and distribute them across teams. To do so, we needed to design a team formation task grounded on a use case that most workers would relate to and understand. We chose an online classroom scenario whereby they were asked to drag and drop profiles of learners showcasing different levels of Openness to Experience, Conscientiousness, and Ability into distinct and equally sized teams. The study was designed to accommodate the following research questions:

RQ4: How does the crowd decide on team formation given profiling attributes?

(a) *Does the even distribution of the team members' attributes differ based on the attribute (i.e., Openness to Experience, Conscientiousness, and Ability)?* This Research has a follow-up sub-question if the answer is true. The sub-question regarding potential disparities in attribute levels, namely high and low, is as follows:

   (a) *Which attribute level is the most evenly distributed?*

   The sub-research Question concerns differences in even distribution between high and low attribute levels.

(b) *Does cohesion differ based on the attribute?*

(c) *Does the team's balance differ based on the attribute?*

To address **RQ4**, the study in Chapter 6 examined the dynamics of team formation in crowdsourcing settings. We asked 102 crowd workers recruited from Prolific to divide a list of individuals (with given attributes) into teams to understand their approach to the team formation problem. The study focused on the influence of individual profile attributes such as Openness to Experience, Conscientiousness, and Ability. The existing literature on the effects of these traits on team performance is vast, as discussed in Chapter 6. For instance, high Openness to Experience is known to foster creativity and pro-activity in teamwork [413]; Conscientious teams are known to excel at performance and structured tasks, especially in self-managed projects [195]; High ability teams are known to contribute significantly to specialized and complex problems [154]. However, we do not yet know whether and how the main stakeholders of crowdsourcing team formation – namely crowd workers – see these traits. In the following paragraphs, we highlight and discuss the main findings from our study.

At least one teammate should have high Openness to Experience. The results from our study reveal a clear preference among crowd workers for evenly distributing all three attributes and, most significantly, Openness to Experience. This implies that each team should include at least one member with a high level of Openness to Experience, ensuring that every team benefits from the unique perspectives and innovative thinking associated with this trait. The figure of a highly Open to Experience teammate can be considered a proxy for what other researchers called *idea scouts*, or *idea connectors* [603]. These are people in an organization adept at identifying and sourcing new ideas and technologies from outside the company. In their work on deconstructing the success of companies that have excelled in market leadership through open innovation (such as Procter & Gamble, Cisco Systems, and Intel), Whelan et al. [603] refers to these figures as instrumental. According to the authors, idea scouts act as the team's antennas, scanning the external environment for innovations and idea connectors between collaborators.

While emphasizing Openness to Experience, it is also essential to provide a balance with other attributes, such as Conscientiousness and Ability, as the study shows a preference for evenly distributing these traits. This balance can help maintain fairness, ensuring no team is left behind. However, while balancing for fairness is a noble objective (since it strives to give all teams equal chances and similar starting points), balancing attributes should also be moderated by other constraints, including the utility of the team formation and the rights and autonomy of the individuals [497]. This means that fairness through balance, while critical, must be considered alongside the preferences and rights of individuals and the objectives of the task – which can sometimes be at odds with balance and equitable treatment of all teams.

High levels of personality traits matter and must be balanced. High attributes were more evenly distributed across teams than low levels, with Openness to Experience again showing the most even distribution among the high levels. We discuss the valence of these findings in light of the study by Carter et al. [81]. Their work, defining the desirability property of different Big-5 personalities, provides evidence that extremely high levels of a given trait can be maladaptive rather than beneficial. For example, excessive Openness to Experience may predispose individuals to less clear differentiation between reality and fantasy, potentially affecting team dynamics and decision-making processes. This suggests that while high traits such as Openness to Experience are valued for bringing diverse perspectives and innovation, an overabundance might not always be beneficial. Balancing these high levels means pairing single individuals with high levels with others with medium and low levels of the same traits. This can also be achieved by moderating across attributes such as Conscientiousness and Ability, which can help maintain a dynamic equilibrium within teams, ensuring that no team is disadvantaged by having members with overly extreme personality profiles. Furthermore, the study by Carter et al. [81] evidences that having (at least) one individual with high Openness to Experience is not enough to guarantee a successful teamwork outcome since intelligence (or, as we called it in the study, Ability) is a necessary condition for Openness to manifest in creative achievement.

Teammates with the same level of Conscientiousness are preferred. The question of whether cohesion differs based on the chosen attribute is also addressed. The study found that while most teams formed were cohesive, the attribute of Conscientiousness most consistently led to high team cohesion. This implies that crowd workers might have thought teams with a low variety of Conscientiousness tend to work more harmoniously, perhaps due to a shared work ethic and attention to detail [193], or the lack thereof. The literature on whether Conscientiousness in teams is directly linked with performance is contradictory and varies according to the task and the team structure (e.g., [508, 484, 581]). However, in our results, we note that the *similarity in Conscientiousness* counts, rather than its high levels. Thus, at least, according to crowd workers, teams whose members share similar attitudes towards work ethics, impulse control, regulation, and direction (some of the most known social aspects influenced by Conscientiousness [200]) are preferred over others where these attitudes are clashing. While our results from the study of teams under pressure (Chapter 5) showed that diverse levels of Conscientiousness produced more positive impressions amongst the teammates, this may not hold for other tasks with different objectives, team dynamics, and communication channels, as also highlighted in the literature [194].

7.6. Limitations

Despite making every effort towards a broad and diverse set of studies, we list some of the most salient limitations to consider when advancing the research on user-centred crowdsourcing teams. These limitations further ground and contextualize the contributions of this thesis.

Chapter 3: Online contexts and survey limitations. The limitations in Chapter 3 revolve around the generalizability of findings from survey data collected via crowdsourcing platforms. This method may not accurately reflect crowd workers' real-world decisions in team formation due to the specific online context, which could influence participants' preferences and behaviours. The reliance on surveys for profiling attributes may not capture the complexities of real-life scenarios where the choice of trait disclosure is made.

Chapter 4: Digital nudging in team formation. In Chapter 4, the study's limitations are tied to its experimental setting involving a fictitious company and fictional teammates. Using digital nudging techniques to influence team diversity may not represent the real-world complexities of team formation. The specific experimental context might not account for the broader range of factors that affect the selection of teammates and the formation of diverse teams in professional or collaborative environments. Additionally, the study's approach to measuring and interpreting behaviour through controlled experiments may not capture the spontaneity and complexity of human interactions in natural settings.

Chapter 5: Virtual and simulated environments. Chapter 5's study, utilizing a virtual maze game to simulate high-pressure scenarios like bomb defusal, might not accurately reflect real-world emergencies. The virtual setting could affect participants' behaviours and decisions differently than in real-life crises. The simplification necessary for the study's execution and reliance on pre- and post-task surveys and quantitative metrics like task completion time and error rates may not capture the qualitative aspects of team dynamics and decision-making processes in emergency responses. Additionally, the simplification inherent in using tools like the Big-5 Inventory-10 for personality measurement may not fully encompass the nuances of human personality and behaviour in various contexts. The study's focus on self-reported data also raises concerns about biases like social desirability, potentially distorting true preferences.

Chapter 6: Controlled experimental design. While ideal for research, the controlled experimental design of Chapter 6 lacks the realism of actual team formation scenarios. The User as Wizard approach might not encompass real-world team assemblies' complexities and dynamic interactions, such as in professional or educational environments. This limitation suggests that, although the study yields valuable insights into team formation preferences, it might not cover all aspects of team dynamics and individual behaviours across various real-life contexts.

7.7. Future Work

In light of the contribution to knowledge and limitations provided by the thesis' studies, we suggest the following list of future work.

Deepening understanding of profiling preferences. Following the work proposed in Chapter 3, further research should explore the nuances of crowd workers' preferences for surface- and deep-level traits across diverse platforms and cultural contexts. Crowdworkers profiling themselves for software engineering collaborative projects such as those from Kaggle [54] may yield different outcomes than other types of contests and contexts (e.g., creative solutions for wicked problems as shown on OpenIDEO [316]). Such an investigation would enrich the understanding of global dynamics in online team formation and improve the contextualization of the disclosure of crowd workers' traits.

Exploring diverse digital nudging techniques. Since inclusion, diversity, and human-centred design are core aspects of this thesis, we deem it essential to continue researching how digital interventions can encourage crowd workers towards more inclusive and diverse decisions. Following the work done in Chapter 4, we suggest investigating various digital nudging strategies, including their long-term impacts on team diversity, in different task scenarios and cultural contexts, preferably relying on real tasks and teammates.

Expanding emergency-response crowdsourcing task research. Future studies should investigate varied stress-inducing tasks, delving into how virtual environments interact with crowd workers' personality traits and demographic factors. Future studies could systematically test various types of emergency response tasks either by expanding the complexity and breadth of the virtual environment in Chapter 5 or by teaming up with established institutes and organizations managing emergencies of different natures (e.g., non-profit organizations, emergency communities in rural areas, charities, and humanitarian foundations).

Broadening the User as Wizard study for team formation. The User as Wizard method used in Chapter 6 to elicit human judgement and decision-making in team formation with limited information should be applied in different online contexts, such as professional workplaces. Incorporating the task into scenarios where the teammates are real and their collaboration is critical to accomplishing goals will help explore how various (real-world) attributes influence team dynamics and outcomes.

7.8. Concluding Remarks

This thesis advocates centring design around the user, venturing beyond task completion. It demonstrates how fostering teamwork in crowdsourcing environments can significantly improve by customizing systems to align with individual users' unique preferences, abilities, and motivations. To achieve this, the research has engaged in an in-depth exploration of crowd workers' opinions and behaviours online concerning disclosure, inclusion, teamwork dynamics, and decision-making. It has employed strategies from User-Centered Design (UCD) and Human-Computer Interaction disciplines, focusing on improving systems responsive to the actual behaviours and needs of the primary stakeholders involved. This aligns with contemporary research counteracting the unidirectionality of crowdsourcing systems where the information architecture and processes are designed for the benefit of the requester only.

The collection of works presented in this thesis highlights the importance of prioritizing the users' needs and behaviour to enhance their choices and chances of finding the right collaborators for the task. Chapter 3 focused on the types of profiling attributes users are comfortable sharing and find useful. Chapter 4's examination of digital nudging techniques has contributed significantly to our understanding of promoting diversity in team assembly. Chapter 5's exploration into stress-induced tasks within a virtual environment has shed light on the importance of personality traits and communication patterns in high-pressure team dynamics. Chapter 6 examined the complex process of team formation, emphasizing the role of participant input and profiling attributes. Collectively, these studies indicate the potential of user-centred design in enhancing the efficacy and inclusivity of crowdsourcing platforms. They pave the way for future research that further investigates and refines these principles, continually adapting to the evolving landscape of online collaboration and teamwork.

8

Appendix

Table 8.1: Results for the Cohesion and Balance metrics for the three attributes plus the frequency count of the combinations of teams.

Openness					Conscientiousness					Ability					teams				#
H	M	L	Coh.	Bal?	H	M	L	Coh.	Bal?	H	M	L	Coh.	Bal?	1	2	3	4	
1	1	1	1.00	Yes	-	2	1	0.58	No	1	1	1	1.00	Yes	13	1	-	-	14
-	2	1	0.58	No	-	2	1	0.58	No	1	1	1	1.00	Yes	1	1	-	-	2
-	2	1	0.58	No	-	1	2	0.58	No	2	1	-	0.58	No	1	-	-	-	1
1	1	1	1.00	Yes	-	1	2	0.58	No	1	2	-	0.58	No	4	-	-	-	4
-	2	1	0.58	No	1	1	1	1.00	Yes	1	2	-	0.58	No	1	-	-	-	1
1	1	1	1.00	Yes	-	3	-	0.00	Yes	1	-	2	1.15	No	5	-	-	-	5
1	1	1	1.00	Yes	-	2	1	0.58	No	2	-	1	1.15	No	3	1	-	-	4
2	-	1	1.15	No	-	2	1	0.58	No	1	1	1	1.00	Yes	2	-	-	-	2
1	1	1	1.00	Yes	1	2	-	0.58	No	1	1	1	1.00	Yes	14	1	-	-	15
1	-	2	1.15	No	-	3	-	0.00	Yes	1	1	1	1.00	Yes	2	-	-	-	2
1	1	1	1.00	Yes	-	3	-	0.00	Yes	2	-	1	1.15	No	2	-	-	-	2
1	-	2	1.15	No	1	2	-	0.58	No	1	1	1	1.00	Yes	3	-	-	-	3
1	1	1	1.00	Yes	1	2	-	0.58	No	1	-	2	1.15	No	1	1	-	1	3
1	1	1	1.00	Yes	-	2	1	0.58	No	1	1	1	1.00	Yes	3	1	-	-	4
-	2	1	0.58	No	-	3	-	0.00	Yes	2	-	1	1.15	No	2	-	-	-	2
-	1	2	0.58	No	1	2	-	0.58	No	1	1	1	1.00	Yes	1	-	-	-	1
1	1	1	1.00	Yes	-	3	-	0.00	Yes	1	1	1	1.00	Yes	-	-	1	-	1
-	2	1	0.58	No	1	2	-	0.58	No	1	-	2	1.15	No	1	-	-	-	1
1	1	1	1.00	Yes	-	1	2	0.58	No	2	1	-	0.58	No	2	-	1	-	3
-	2	1	0.58	No	-	2	1	0.58	No	3	-	-	0.00	No	2	-	-	1	3
-	1	2	0.58	No	1	1	1	1.00	Yes	2	1	-	0.58	No	2	-	-	-	2
1	1	1	1.00	Yes	1	1	1	1.00	Yes	1	2	-	0.58	No	3	1	-	-	4
1	-	2	1.15	No	-	2	1	0.58	No	1	2	-	0.58	No	-	1	1	1	3
1	1	1	1.00	Yes	-	2	1	0.58	No	2	1	-	0.58	No	-	1	-	-	1
1	1	1	1.00	Yes	1	1	1	1.00	Yes	1	1	1	1.00	Yes	1	1	-	1	3
1	1	1	1.00	Yes	1	2	-	0.58	No	1	2	-	0.58	No	1	-	1	-	2
-	2	1	0.58	No	2	1	-	0.58	No	1	1	1	1.00	Yes	1	-	-	-	1
-	1	2	0.58	No	-	3	-	0.00	Yes	2	1	-	0.58	No	-	1	-	-	1
-	-	3	0.00	No	1	2	-	0.58	No	1	2	-	0.58	No	3	-	-	1	4
1	-	2	1.15	No	-	3	-	0.00	Yes	1	2	-	0.58	No	-	-	1	-	1
-	1	2	0.58	No	1	2	-	0.58	No	1	1	1	1.00	Yes	1	-	1	1	3
1	1	1	1.00	Yes	-	3	-	0.00	Yes	2	1	-	0.58	No	-	-	2	1	3
1	-	2	1.15	No	1	2	-	0.58	No	1	2	-	0.58	No	-	-	1	-	1
1	2	-	0.58	No	1	1	1	1.00	Yes	-	2	1	0.58	No	1	-	-	-	1
1	1	1	1.00	Yes	-	2	1	0.58	No	-	2	1	0.58	No	-	1	1	-	2
1	2	-	0.58	No	-	2	1	0.58	No	1	1	1	1.00	Yes	-	1	-	-	1
1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	2	1	0.58	No	2	-	-	-	2
1	2	-	0.58	No	1	1	1	1.00	Yes	-	1	2	0.58	No	1	1	-	-	2
-	3	-	0.00	Yes	-	1	2	0.58	No	1	1	1	1.00	Yes	1	-	-	-	1
1	2	-	0.58	No	-	1	2	0.58	No	-	2	1	0.58	No	-	-	1	-	1
-	3	-	0.00	Yes	1	1	1	1.00	Yes	-	2	1	0.58	No	1	-	1	1	3
-	2	1	0.58	No	-	2	1	0.58	No	-	2	1	0.58	No	-	2	1	2	5
-	3	-	0.00	Yes	-	2	1	0.58	No	1	1	1	1.00	Yes	-	-	1	2	3
-	2	1	0.58	No	1	1	1	1.00	Yes	-	2	1	0.58	No	1	-	1	1	3
1	2	-	0.58	No	-	2	1	0.58	No	-	2	1	0.58	No	-	1	-	-	1
1	2	-	0.58	No	-	-	3	0.00	No	1	2	-	0.58	No	-	3	-	-	3
-	3	-	0.00	Yes	1	-	2	1.15	No	1	2	-	0.58	No	1	4	-	-	5
-	2	1	0.58	No	-	1	2	0.58	No	1	2	-	0.58	No	1	-	-	-	1
-	3	-	0.00	Yes	-	1	2	0.58	No	2	1	-	0.58	No	-	-	2	-	2
-	2	1	0.58	No	1	-	2	1.15	No	1	2	-	0.58	No	-	1	-	1	2
1	2	-	0.58	No	1	-	2	1.15	No	-	3	-	0.00	Yes	1	1	-	1	3
1	1	1	1.00	Yes	-	1	2	0.58	No	-	3	-	0.00	Yes	1	-	-	-	1
1	2	-	0.58	No	-	1	2	0.58	No	1	2	-	0.58	No	-	-	1	-	1
1	1	1	1.00	Yes	1	-	2	1.15	No	-	3	-	0.00	Yes	-	1	1	1	3
2	1	-	0.58	No	-	1	2	0.58	No	-	3	-	0.00	Yes	-	-	1	-	1
1	2	-	0.58	No	1	-	2	1.15	No	-	2	1	0.58	No	-	2	-	-	2
-	2	1	0.58	No	1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	2	2	-	4
-	3	-	0.00	Yes	1	1	1	1.00	Yes	1	2	-	0.58	No	-	-	-	2	2
1	2	-	0.58	No	1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	-	1	1	2
-	3	-	0.00	Yes	2	-	1	1.15	No	-	2	1	0.58	No	-	-	1	-	1
1	1	1	1.00	Yes	-	2	1	0.58	No	-	3	-	0.00	Yes	-	-	1	2	3
-	2	1	0.58	No	1	1	1	1.00	Yes	-	2	1	0.58	No	-	-	-	1	1
-	2	1	0.58	No	1	1	1	1.00	Yes	1	2	-	0.58	No	-	1	-	1	2

Table 8.1: Results for the Cohesion and Balance metrics for the three attributes plus the frequency count of the combinations of teams.

Openness					Conscientiousness					Ability					teams				#
H	M	L	Coh.	Bal?	H	M	L	Coh.	Bal?	H	M	L	Coh.	Bal?	1	2	3	4	
1	2	-	0.58	No	-	2	1	0.58	No	1	2	-	0.58	No	-	1	1	-	2
-	3	-	0.00	Yes	1	1	1	1.00	Yes	1	1	1	1.00	Yes	1	-	3	-	4
1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	-	3	1	4
-	2	1	0.58	No	2	-	1	1.15	No	-	2	1	0.58	No	-	1	-	1	2
1	2	-	0.58	No	1	1	1	1.00	Yes	-	2	1	0.58	No	-	-	-	4	4
1	2	-	0.58	No	-	2	1	0.58	No	1	-	2	1.15	No	1	1	-	-	2
2	1	-	0.58	No	-	2	1	0.58	No	-	1	2	0.58	No	-	1	-	-	1
1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	1	2	0.58	No	-	1	1	-	2
1	2	-	0.58	No	-	3	-	0.00	Yes	1	-	2	1.15	No	1	-	-	-	1
1	1	1	1.00	Yes	1	2	-	0.58	No	-	1	2	0.58	No	1	-	-	-	1
2	1	-	0.58	No	-	3	-	0.00	Yes	-	1	2	0.58	No	-	1	-	-	1
1	2	-	0.58	No	1	2	-	0.58	No	-	-	3	0.00	No	-	1	1	-	2
1	2	-	0.58	No	1	1	1	1.00	Yes	1	1	1	1.00	Yes	2	2	-	-	4
1	1	1	1.00	Yes	-	2	1	0.58	No	1	1	1	1.00	Yes	-	1	1	-	2
1	1	1	1.00	Yes	1	1	1	1.00	Yes	1	1	1	1.00	Yes	2	2	1	-	5
2	1	-	0.58	No	-	2	1	0.58	No	1	1	1	1.00	Yes	-	-	-	1	1
1	2	-	0.58	No	1	1	1	1.00	Yes	1	-	2	1.15	No	-	2	-	-	2
2	1	-	0.58	No	1	1	1	1.00	Yes	-	2	1	0.58	No	-	1	-	-	1
3	-	-	0.00	No	-	2	1	0.58	No	-	2	1	0.58	No	1	2	-	-	3
2	1	-	0.58	No	1	1	1	1.00	Yes	-	1	2	0.58	No	-	1	-	-	1
1	1	1	1.00	Yes	2	1	-	0.58	No	-	2	1	0.58	No	-	-	-	1	1
1	1	1	1.00	Yes	-	3	-	0.00	Yes	1	1	1	1.00	Yes	-	2	-	-	2
1	-	2	1.15	No	1	2	-	0.58	No	-	2	1	0.58	No	-	1	-	1	2
2	-	1	1.15	No	-	3	-	0.00	Yes	-	2	1	0.58	No	-	-	-	1	1
1	1	1	1.00	Yes	1	2	-	0.58	No	-	1	2	0.58	No	1	1	1	-	3
1	1	1	1.00	Yes	1	2	-	0.58	No	1	1	1	1.00	Yes	-	-	2	-	2
1	2	-	0.58	No	1	2	-	0.58	No	1	-	2	1.15	No	-	-	1	1	2
1	1	1	1.00	Yes	2	1	-	0.58	No	-	1	2	0.58	No	-	-	1	-	1
2	1	-	0.58	No	1	2	-	0.58	No	-	1	2	0.58	No	-	-	1	-	1
1	2	-	0.58	No	-	1	2	0.58	No	1	1	1	1.00	Yes	-	2	2	-	4
-	3	-	0.00	Yes	1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	5	-	1	6
-	2	1	0.58	No	-	2	1	0.58	No	1	1	1	1.00	Yes	-	1	-	-	1
-	3	-	0.00	Yes	-	2	1	0.58	No	2	-	1	1.15	No	-	2	-	-	2
-	2	1	0.58	No	1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	1	-	1	2
1	2	-	0.58	No	-	2	1	0.58	No	1	1	1	1.00	Yes	-	1	1	-	2
-	3	-	0.00	Yes	1	1	1	1.00	Yes	1	-	2	1.15	No	-	1	3	1	5
1	2	-	0.58	No	1	1	1	1.00	Yes	-	2	1	0.58	No	-	-	1	-	1
1	1	1	1.00	Yes	-	2	1	0.58	No	-	2	1	0.58	No	-	1	-	-	1
1	2	-	0.58	No	-	2	1	0.58	No	1	1	1	1.00	Yes	-	3	1	1	5
1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	2	1	0.58	No	-	1	-	-	1
-	2	1	0.58	No	1	2	-	0.58	No	-	2	1	0.58	No	-	1	-	1	2
-	3	-	0.00	Yes	1	2	-	0.58	No	1	1	1	1.00	Yes	-	-	2	1	3
-	3	-	0.00	Yes	2	1	-	0.58	No	-	1	2	0.58	No	-	-	1	1	2
-	1	2	0.58	No	1	2	-	0.58	No	-	2	1	0.58	No	-	-	1	-	1
1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	2	1	0.58	No	-	-	-	3	3
-	2	1	0.58	No	1	2	-	0.58	No	-	1	2	0.58	No	-	-	-	2	2
-	2	1	0.58	No	1	2	-	0.58	No	1	1	1	1.00	Yes	-	-	1	1	2
1	2	-	0.58	No	-	3	-	0.00	Yes	1	1	1	1.00	Yes	1	1	1	2	5
-	3	-	0.00	Yes	1	2	-	0.58	No	1	-	2	1.15	No	1	-	-	2	3
1	1	1	1.00	Yes	1	2	-	0.58	No	-	2	1	0.58	No	-	2	1	3	6
1	2	-	0.58	No	1	-	2	1.15	No	1	2	-	0.58	No	-	4	1	-	5
1	1	1	1.00	Yes	1	-	2	1.15	No	1	2	-	0.58	No	1	3	1	1	6
1	2	-	0.58	No	1	-	2	1.15	No	1	1	1	1.00	Yes	-	-	1	3	4
-	2	1	0.58	No	1	1	1	1.00	Yes	1	2	-	0.58	No	-	1	1	-	2
-	2	1	0.58	No	2	-	1	1.15	No	1	2	-	0.58	No	-	-	2	-	2
1	2	-	0.58	No	1	1	1	1.00	Yes	1	2	-	0.58	No	-	-	2	-	2
-	2	1	0.58	No	-	2	1	0.58	No	2	1	-	0.58	No	-	1	-	-	1
1	1	1	1.00	Yes	-	2	1	0.58	No	1	2	-	0.58	No	-	-	-	1	1
-	2	1	0.58	No	1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	-	1	1	2
-	2	1	0.58	No	1	1	1	1.00	Yes	2	1	-	0.58	No	-	-	1	2	3
-	3	-	0.00	Yes	1	1	1	1.00	Yes	2	-	1	1.15	No	-	1	-	1	2
1	1	1	1.00	Yes	1	1	1	1.00	Yes	1	2	-	0.58	No	-	2	1	2	5
-	2	1	0.58	No	2	-	1	1.15	No	1	1	1	1.00	Yes	-	-	-	2	2

Table 8.1: Results for the Cohesion and Balance metrics for the three attributes plus the frequency count of the combinations of teams.

Openness					Conscientiousness					Ability					teams				#
H	M	L	Coh.	Bal?	H	M	L	Coh.	Bal?	H	M	L	Coh.	Bal?	1	2	3	4	
1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	-	1	-	1
1	2	-	0.58	No	1	1	1	1.00	Yes	1	2	-	0.58	No	-	1	-	-	1
1	1	1	1.00	Yes	2	-	1	1.15	No	-	3	-	0.00	Yes	-	-	2	-	2
2	1	-	0.58	No	1	1	1	1.00	Yes	-	3	-	0.00	Yes	1	-	-	1	2
1	1	1	1.00	Yes	-	2	1	0.58	No	1	2	-	0.58	No	-	-	-	2	2
1	-	2	1.15	No	1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	-	2	-	2
1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	2	1	0.58	No	1	4	2	1	8
1	1	1	1.00	Yes	1	1	1	1.00	Yes	1	2	-	0.58	No	-	2	2	-	4
2	1	-	0.58	No	-	2	1	0.58	No	1	2	-	0.58	No	-	-	1	-	1
1	2	-	0.58	No	1	1	1	1.00	Yes	1	1	1	1.00	Yes	-	1	-	3	4
2	-	1	1.15	No	1	1	1	1.00	Yes	-	3	-	0.00	Yes	-	1	-	-	1
1	1	1	1.00	Yes	2	-	1	1.15	No	-	2	1	0.58	No	-	-	-	1	1
2	1	-	0.58	No	1	1	1	1.00	Yes	-	2	1	0.58	No	-	-	1	-	1
-	2	1	0.58	No	1	2	-	0.58	No	1	2	-	0.58	No	-	-	1	1	2
-	1	2	0.58	No	2	1	-	0.58	No	-	3	-	0.00	Yes	-	-	1	1	2
1	1	1	1.00	Yes	1	2	-	0.58	No	-	3	-	0.00	Yes	-	-	-	2	2
-	2	1	0.58	No	2	1	-	0.58	No	-	2	1	0.58	No	-	-	1	1	2
1	2	-	0.58	No	1	2	-	0.58	No	1	2	-	0.58	No	-	1	3	3	7
-	3	-	0.00	Yes	2	1	-	0.58	No	1	1	1	1.00	Yes	-	-	1	1	2
-	2	1	0.58	No	3	-	-	0.00	No	-	2	1	0.58	No	1	-	2	-	3
1	2	-	0.58	No	2	1	-	0.58	No	-	2	1	0.58	No	-	-	-	1	1
1	1	1	1.00	Yes	-	3	-	0.00	Yes	1	2	-	0.58	No	-	3	1	1	5
-	2	1	0.58	No	1	2	-	0.58	No	1	1	1	1.00	Yes	-	-	2	-	2
1	-	2	1.15	No	1	2	-	0.58	No	-	3	-	0.00	Yes	-	-	4	2	6
-	1	2	0.58	No	2	1	-	0.58	No	-	2	1	0.58	No	-	-	-	2	2
1	1	1	1.00	Yes	1	2	-	0.58	No	-	2	1	0.58	No	-	-	2	5	7
1	1	1	1.00	Yes	1	2	-	0.58	No	1	2	-	0.58	No	-	-	4	1	5
1	2	-	0.58	No	1	2	-	0.58	No	1	1	1	1.00	Yes	-	1	-	4	5
1	1	1	1.00	Yes	2	1	-	0.58	No	-	2	1	0.58	No	-	-	1	2	3

Openness to Experience		**Conscientiousness**		**Ability**		
Open AvgTA	**Class**	**Cons AvgTA**	**Class**	**Able AvgTA**	**Class**	**Total**
2.00	M	1.67	M	2.00	M	14
1.67	M	1.67	M	2.00	M	2
1.67	M	1.33	L	2.67	H	1
2.00	M	1.33	L	2.33	H	4
1.67	M	2.00	M	2.33	H	1
2.00	M	2.00	M	1.67	M	5
2.00	M	1.67	M	2.33	H	4
2.33	H	1.67	M	2.00	M	2
2.00	M	2.33	H	2.00	M	15
1.67	M	2.00	M	2.00	M	2
2.00	M	2.00	M	2.33	H	2
1.67	M	2.33	H	2.00	M	3
2.00	M	2.33	H	1.67	M	3
2.00	M	1.67	M	2.00	M	4
1.67	M	2.00	M	2.33	H	2
1.33	L	2.33	H	2.00	M	1
2.00	M	2.00	M	2.00	M	1
1.67	M	2.33	H	1.67	M	1
2.00	M	1.33	L	2.67	H	3
1.67	M	1.67	M	3.00	H	3
1.33	L	2.00	M	2.67	H	2
2.00	M	2.00	M	2.33	H	4
1.67	M	1.67	M	2.33	H	3
2.00	M	1.67	M	2.67	H	1
2.00	M	2.00	M	2.00	M	3
2.00	M	2.33	H	2.33	H	2
1.67	M	2.67	H	2.00	M	1

Openness to Experience		Conscientiousness		Ability		
Open AvgTA	**Class**	**Cons AvgTA**	**Class**	**Able AvgTA**	**Class**	**Total**
1.33	L	2.00	M	2.67	H	1
1.00	L	2.33	H	2.33	H	4
1.67	M	2.00	M	2.33	H	1
1.33	L	2.33	H	2.00	M	3
2.00	M	2.00	M	2.67	H	3
1.67	M	2.33	H	2.33	H	1
2.33	H	2.00	M	1.67	M	1
2.00	M	1.67	M	1.67	M	2
2.33	H	1.67	M	2.00	M	1
2.00	M	2.00	M	1.67	M	2
2.33	H	2.00	M	1.33	L	2
2.00	M	1.33	L	2.00	M	1
2.33	H	1.33	L	1.67	M	1
2.00	M	2.00	M	1.67	M	3
1.67	M	1.67	M	1.67	M	5
2.00	M	1.67	M	2.00	M	3
1.67	M	2.00	M	1.67	M	3
2.33	H	1.67	M	1.67	M	1
2.33	H	1.00	L	2.33	H	3
2.00	M	1.67	M	2.33	H	5
1.67	M	1.33	L	2.33	H	1
2.00	M	1.33	L	2.67	H	2
1.67	M	1.67	M	2.33	H	2
2.33	H	1.67	M	2.00	M	3
2.00	M	1.33	L	2.00	M	1
2.33	H	1.33	L	2.33	H	1
2.00	M	1.67	M	2.00	M	3
2.67	H	1.33	L	2.00	M	1
2.33	H	1.67	M	1.67	M	2
1.67	M	2.00	M	2.00	M	4
2.00	M	2.00	M	2.33	H	2
2.33	H	2.00	M	2.00	M	2
2.00	M	2.33	H	1.67	M	1
2.00	M	1.67	M	2.00	M	3
1.67	M	2.00	M	1.67	M	1
1.67	M	2.00	M	2.33	H	2
2.33	H	1.67	M	2.33	H	2
2.00	M	2.00	M	2.00	M	4
2.00	M	2.00	M	2.00	M	4
1.67	M	2.33	H	1.67	M	2
2.33	H	2.00	M	1.67	M	4
2.33	H	1.67	M	1.67	M	2
2.67	H	1.67	M	1.33	L	1
2.00	M	2.00	M	1.33	L	2
2.33	H	2.00	M	1.67	M	1
2.00	M	2.33	H	1.33	L	1
2.67	H	2.00	M	1.33	L	1
2.33	H	2.33	H	1.00	L	2
2.33	H	2.00	M	2.00	M	4
2.00	M	1.67	M	2.00	M	2
2.00	M	2.00	M	2.00	M	5
2.67	H	1.67	M	2.00	M	1
2.33	H	2.00	M	1.67	M	2
2.67	H	2.00	M	1.67	M	1
3.00	H	1.67	M	1.67	M	3
2.67	H	2.00	M	1.33	L	1
2.00	M	2.67	H	1.67	M	1
2.00	M	2.00	M	2.00	M	2
1.67	M	2.33	H	1.67	M	2
2.33	H	2.00	M	1.67	M	1
2.00	M	2.33	H	1.33	L	3
2.00	M	2.33	H	2.00	M	2
2.33	H	2.33	H	1.67	M	2
2.00	M	2.67	H	1.33	L	1
2.67	H	2.33	H	1.33	L	1
2.33	H	1.33	L	2.00	M	4
2.00	M	2.00	M	2.00	M	6

Openness to Experience		Conscientiousness		Ability		
Open AvgTA	**Class**	**Cons AvgTA**	**Class**	**Able AvgTA**	**Class**	**Total**
1.67	M	1.67	M	2.00	M	1
2.00	M	1.67	M	2.33	H	2
1.67	M	2.00	M	2.00	M	2
2.33	H	1.67	M	2.00	M	2
2.00	M	2.00	M	1.67	M	5
2.33	H	2.00	M	1.67	M	1
2.00	M	1.67	M	1.67	M	1
2.33	H	1.67	M	2.00	M	5
2.00	M	2.00	M	1.67	M	1
1.67	M	2.33	H	1.67	M	2
2.00	M	2.33	H	2.00	M	3
2.00	M	2.67	H	1.33	L	2
1.33	L	2.33	H	1.67	M	1
2.00	M	2.00	M	1.67	M	3
1.67	M	2.33	H	1.33	L	2
1.67	M	2.33	H	2.00	M	2
2.33	H	2.00	M	2.00	M	5
2.00	M	2.33	H	1.67	M	3
2.00	M	2.33	H	1.67	M	6
2.33	H	1.67	M	2.33	H	5
2.00	M	1.67	M	2.33	H	6
2.33	H	1.67	M	2.00	M	4
1.67	M	2.00	M	2.33	H	2
1.67	M	2.33	H	2.33	H	2
2.33	H	2.00	M	2.33	H	2
1.67	M	1.67	M	2.67	H	1
2.00	M	1.67	M	2.33	H	1
1.67	M	2.00	M	2.00	M	2
1.67	M	2.00	M	2.67	H	3
2.00	M	2.00	M	2.33	H	2
2.00	M	2.00	M	2.33	H	5
1.67	M	2.33	H	2.00	M	2
2.00	M	2.00	M	2.00	M	1
2.33	H	2.00	M	2.33	H	1
2.00	M	2.33	H	2.00	M	2
2.67	H	2.00	M	2.00	M	2
2.00	M	1.67	M	2.33	H	2
1.67	M	2.00	M	2.00	M	2
2.00	M	2.00	M	1.67	M	8
2.00	M	2.00	M	2.33	H	4
2.67	H	1.67	M	2.33	H	1
2.33	H	2.00	M	2.00	M	4
2.33	H	2.00	M	2.00	M	1
2.00	M	2.33	H	1.67	M	1
2.67	H	2.00	M	1.67	M	1
1.67	M	2.33	H	2.33	H	2
1.33	L	2.67	H	2.00	M	2
2.00	M	2.33	H	2.00	M	2
1.67	M	2.67	H	1.67	M	2
2.33	H	2.33	H	2.33	H	7
2.00	M	2.67	H	2.00	M	2
1.67	M	3.00	H	1.67	M	3
2.33	H	2.67	H	1.67	M	1
2.00	M	2.00	M	2.33	H	5
1.67	M	2.33	H	2.00	M	2
1.67	M	2.33	H	2.00	M	6
1.33	L	2.67	H	1.67	M	2
2.00	M	2.33	H	1.67	M	7
2.00	M	2.33	H	2.33	H	5
2.33	H	2.33	H	2.00	M	5
2.00	M	2.67	H	1.67	M	3
Openness to experience		Conscientiousness		Ability		Total TA
LTA	16	LTA	23	LTA	19	#LTA: 58
MTA	292	MTA	254	MTA	269	#MTA: 815
HTA	100	HTA	131	HTA	120	#HTA: 351

Table 8.2: Informed consent as presented to the participants before the task.

Informed consent
You are invited to participate in a study about finding a new coffee slogan for Xpresso. This study aims to gain insight into how digital interfaces influence the choice of team members in professional contexts. This research is conducted for a bachelor thesis at Utrecht University. It should take about 5-8 minutes to complete.
Voluntary participation
Your participation is voluntary. You may refuse to take part in the task or exit the task at any time. However, to receive the reward, you need to have completed the task.
Privacy and confidentiality
All the collected data, including information from the profile you will create, will be anonymized and not shared beyond the study team (which consists of a bachelor student and two supervisors). Your privacy is protected. The username of your profile will only be used for the experiment and deleted after the experiment is completed.
I've had the opportunity to read this informed consent page, and I agree to participate in this study.
[tick box] I consent

Table 8.3: Distribution of diversity attributes among dummy profiles (30 dummies)

Attribute	**Category**	**Dummies within category**
BACKGROUND	Information systems	3
	Customer service	3
	Sales and marketing	3
	Creative sector	3
	Engineering, R&D	3
	Purchasing/procurement	3
	Operations, administrations, manufacturing	3
	General management	3
	HR/personnel	3
	Consultancy	3
AGE	Generation Z	9
	Milennials	10
	Generation X	11
	Boomers	0
	Greatest generation	0
REGION	European	7
	North African, Middle Eastern or Central Asian	1
	Latin American	3
	East Asian	3
	South and South-East Asian	7
	Carribean	1
	Sub-Saharan African	1
	North-American and Australasian	7
ETHINICITY	white	12
	black	4
	asian	11
	latino	3
GENDER	male	14
	female	15
	other	1

Table 8.4: Task description is shown to all participants before commencing the experiment.

Your task
We are Xpresso, a coffee company, and we are looking for a new company slogan. We need fresh ideas, so we decided to outsource this project. Your task is to select two team members from a list of previously registered individuals to form a team with whom you will collaborate on this project. Out of the teams that are formed by people like you, we choose one team which we think works best. If you choose the best team, you get an additional bonus. Then, you might be contacted to create a slogan with the team you chose. You are free to decline the follow-up task of creating a slogan with the team you chose. It's highly preferred to complete the task on a tablet or computer.
The task consists of four parts.
(1) First, you create a profile profile by registering as a 'new user by adding personal information to your profile.
(2) Then, you'll read some basic information about the company to learn more about Xpresso. It is important you read this information carefully. We will ask you some questions about the content.
(3) Then, you will have to select 2 team members with whom you'd prefer to work with.
(4)Finally, after responding to 10 statements you've completed the task.
[Next button] Start task

Table 8.5: Registration form as seen from the participants.

Welcome. Register and get started.

Registration form	**Option**
Username	*Text box*
Gender	*Drop-down*
Highest completed education	*Drop-down*
Background knowledge and skills in...	*Drop-down*
Age	*Drop-down*
Zone	*Drop-down*
Region	*Drop-down*
Ethnicity (which suits you best)	*Drop-down*
I'm completing this task on my phone	*Yes/No*
Password	*Text box*
Repeat Password	*Text box*
Submit	*Button*

Table 8.6: Manipulation check example with corresponding questions. The form was designed to test the understanding and retention of important information about outsourcing company Xpresso.

Manipulation check	
Please answer the following questions.	*Header*
If you don't know the answers, return to the Xpresso information.	*Subheader*
What is the name of the CEO of Xpresso?	*Drop-down*
When was the company founded?	*Drop-down*
Continue to the next step	*Button*

Table 8.7: Ethnic groups: general and specific categorization from the European Standard Classification of Cultural and Ethnic Groups [513].

General regions	Specific zones
European	West European
	North European (Nordic)
	South European
	South-East European
	East European
North African, Middle Eastern and Central Asian	North African, Middle Eastern and Central Asian
	Arab
	Jewish
	Turkish
	Iranian and Central Asian
	Other North African and Middle Eastern
Sub-Saharan Africa	Sub-Saharan African
	West and Central Asian
	Africa's Horn
	East and South African
South and South-East Asian	South and South-East Asian
	South Asian
	Mainland and Buddhist South-East Asian
	Maritime and Muslim South-East Asian
East Asian	Chinese Asian
	North-East Asian
Latin American	South American
	Central American
Carribean	English-speaking Carribean
Non-English speaking Carribean	
North American and Australasian	North American
	North American
	Australasian

References

[1] Madiha Said Abdul-Razik, Abdullah Mubark Kaity, Nawal Saeed Banafaa, and Ghada Wahby El-Hady. Disaster response in a civil war: Lessons on local hospitals capacity. the case of yemen. *International journal of healthcare management*, 14 (1):99–106, 2021.

[2] Olga Abramova. No matter what the name, we're all the same? examining ethnic online discrimination in ridesharing marketplaces. *Electronic Markets*, 32(3): 1419–1446, 2022.

[3] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.

[4] Mervat Abu-Elkheir, Hossam S Hassanein, and Sharief MA Oteafy. Enhancing emergency response systems through leveraging crowdsensing and heterogeneous data. In *2016 international wireless communications and mobile computing conference (IWCMC)*, pages 188–193. IEEE, 2016.

[5] Oguz A Acar. Why crowdsourcing often leads to bad ideas. *Harvard Business Review*, 2019.

[6] Alessandro Acquisti and Christina Fong. An experiment in hiring discrimination via online social networks. *Management Science*, 66(3):1005–1024, 2020.

[7] Eleni Adamopoulou and Lefteris Moussiades. An overview of chatbot technology. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 373–383. Springer, 2020.

[8] Allan Afuah and Christopher L Tucci. Crowdsourcing as a solution to distant search. *Academy of management review*, 37(3):355–375, 2012.

[9] Rakesh Agrawal, Behzad Golshan, and Evimaria Terzi. Grouping students in educational settings. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1017–1026, 2014.

[10] Samir Akachi and Arbia Ayed. Crowd-solving: Trust and motivation. *International Journal of Advanced Engineering and Management Research*, 2(4):1049–1061, 2017.

[11] Harini Alagarai Sampath, Rajeev Rajeshuni, and Bipin Indurkhya. Cognitively inspired task design to improve user performance on crowdsourcing platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3665–3674, 2014.

[12] David Alexander. Towards the development of standards in emergency management training and education. *Disaster Prevention and Management: An International Journal*, 2003.

[13] Abdulwhab Alkharashi and Karen Renaud. Privacy in crowdsourcing: a systematic

review. In *Information Security: 21st International Conference, ISC 2018, Guildford, UK, September 9–12, 2018, Proceedings 21*, pages 387–400. Springer, 2018.

[14] Brenda J Allen and Kavita Garg. Diversity matters in academic radiology: acknowledging and addressing unconscious bias. *Journal of the American College of Radiology*, 13(12):1426–1432, 2016.

[15] David W Allen. *Getting to Know ArcGIS: Modelbuilder*, volume 380. Esri Press Redlands, 2011.

[16] Greg Allen. Understanding ai technology. Technical report, Joint Artificial Intelligence Center (JAIC) The Pentagon United States, 2020.

[17] Sofiane Amara, Joaquim Macedo, Fatima Bendella, and Alexandre Santos. Group formation in mobile computer supported collaborative learning contexts: A systematic literature review. *Journal of Educational Technology & Society*, 19 (2):258–273, 2016.

[18] Joshua Amaya, Y Edward Wen, Zhiguo Shang, Andrew Jamieson, and Al Aly. A crowdsourced evaluation of facial averageness and attractiveness. *Aesthetic surgery journal*, 2022.

[19] Yair Amichai-Hamburger and Gideon Vinitzky. Social network use and personality. *Computers in human behavior*, 26(6):1289–1295, 2010.

[20] Deborah Gladstein Ancona and David F Caldwell. Demography and design: Predictors of new product team performance. *Organization science*, 3(3):321–341, 1992.

[21] Mary Anderson. *How the Help Desk Can Support the Security Team*. PhD thesis, Utica College, 2020.

[22] Varvara Antoniou, Emmanuel Vassilakis, and Maria Hatzaki. Is crowdsourcing a reliable method for mass data acquisition? the case of covid-19 spread in greece during spring 2020. *ISPRS International Journal of Geo-Information*, 9(10):605, 2020.

[23] Jose Apesteguia, Ghazala Azmat, and Nagore Iriberri. The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58:78–93, 2012.

[24] Ofer Arazy, Eleni Stroulia, Stan Ruecker, Cristina Arias, Carlos Fiorentino, Veselin Ganev, and Timothy Yau. Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. *Journal of the American Society for Information Science and Technology*, 61(6):1166–1179, 2010.

[25] R. C. Archibold. A nation at war: The new organizations; papers debate use of teams to react quickly to terror. *New York Time*, mar 2003.

[26] Aitor Aritzeta, Sabino Ayestaran, and Stephen Swailes. Team role preference and conflict management styles. *International Journal of conflict management*, 2005.

[27] Payal Arora and Linnea Holter Thompson. Crowdsourcing as a platform for digital labor unions. *International Journal of Communication*, 12:2314–2332, 2019.

[28] Daniel Ashton. Producing participatory media:(crowd) sourcing content in britain/life in a day. *Media International Australia*, 154:101–111, 2015.

[29] Michael C Ashton and Kibeom Lee. Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and social psychology review*, 11(2):150–166, 2007.

[30] Gregory Asmolov. Crowdsourcing and the folksonomy of emergency response: The construction of a mediated subject. *Interactions: Studies in Communication & Culture*, 6(2):155–178, 2015.

[31] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. Personality and patterns of facebook usage. In *Proceedings of the 4th annual ACM web science conference*, pages 24–32, 2012.

[32] Judith Bakker, Bas Denters, Mirjan Oude Vrielink, and Pieter-Jan Klok. Citizens' initiatives: How local governments fill their facilitative role. *Local government studies*, 38(4):395–414, 2012.

[33] Thilak R Balasubramanian and Trilce Estrada. Crowdlearning: A framework for collaborative and personalized learning. In *2016 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE, 2016.

[34] Steve Baldwin and Peter A Woods. Case management and needs assessment: Some issues of concern for the caring professions. *Journal of Mental Health*, 3(3): 311–322, 1994.

[35] Sandra Bammert, Ulrich Matthias König, Maximilian Roeglinger, and Tabitha Wruck. Exploring potentials of digital nudging for business processes. *Business Process Management Journal*, 26(6):1329–1347, 2020.

[36] Maya Bar-Hillel et al. Location, location, location: Position effects in choice among simultaneously presented options. Technical report, The Federmann Center for the Study of Rationality, the Hebrew University, Jerusalem., 2011.

[37] Natã M Barbosa and Monchu Chen. Rehumanized crowdsourcing: A labeling framework addressing bias and ethics in machine learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[38] John A Bargh, Mark Chen, and Lara Burrows. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*, 71(2):230, 1996.

[39] Murray R Barrick, Greg L Stewart, Mitchell J Neubert, and Michael K Mount. Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology*, 83(3):377, 1998.

[40] Paul T Bartone and Amy B Adler. Cohesion over time in a peacekeeping medical task force. *Military Psychology*, 11:85–107, 1999.

[41] Dave Bartram. Scalar equivalence of opq32: Big five profiles of 31 countries. *Journal of Cross-Cultural Psychology*, 44:61–83, 2013.

[42] Raiyan Abdul Baten, Richard N Aslin, Gourab Ghoshal, and Ehsan Hoque. Cues to gender and racial identity reduce creativity in diverse social networks. *Scientific reports*, 11(1):10261, 2021.

[43] Gopi Battineni, Nalini Chintalapudi, and Francesco Amenta. Ai chatbot design during an epidemic like the novel coronavirus, 2020.

[44] Daniel J Beal, Robin R Cohen, Michael J Burke, and Christy L McLendon. Cohesion and performance in groups: a meta-analytic clarification of construct relations. *Journal of applied psychology*, 88(6):989, 2003.

[45] R Meredith Belbin. *Team roles at work*. Routledge, 2012.

[46] Suzanne T Bell, David M Fisher, Shanique G Brown, and Kristin E Mann. An approach for conducting actionable research with extreme teams. *Journal of Management*, 44(7):2740–2765, 2018.

[47] Jean-Michel Benkert and Nick Netzer. Informational requirements of nudging. *Journal of Political Economy*, 126(6):2323–2355, 2018.

[48] Mark Bergen. Google outage reignites worries about smart home without backups, dec 2020. URL `https://www.bloomberg.com/news/newsletters/2020-12-16/google-outage-reignites-worries-about-smart-home-without-backups`.

[49] Kalwant Bhopal. 'we can talk the talk, but we're not allowed to walk the walk': the role of equality and diversity staff in higher education institutions in england. *Higher Education*, 85(2):325–339, 2023.

[50] Benedikte Bjerge, Nathan Clark, Peter Fisker, and Emmanuel Raju. Technology and information sharing in disaster relief. *PLoS One*, 11(9):e0161783, 2016.

[51] Irene V Blair, Jennifer E Ma, and Alison P Lenton. Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of personality and social psychology*, 81(5):828, 2001.

[52] Peter Michael Blau. *Inequality and heterogeneity: A primitive theory of social structure*, volume 7. Free Press New York, 1977.

[53] Eric Bogert, Aaron Schecter, and Richard T Watson. Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific reports*, 11:1–9, 2021.

[54] Casper Solheim Bojer and Jens Peder Meldgaard. Kaggle forecasting competitions: An overlooked learning opportunity. *International Journal of Forecasting*, 37(2): 587–603, 2021.

[55] Kenneth A Bollen and Rick H Hoyle. Perceived cohesion: A conceptual and empirical examination. *Social forces*, 69(2):479–504, 1990.

[56] Gary Bolton, Eugen Dimant, and Ulrich Schmidt. When a nudge backfires: Combining (im) plausible deniability with social and economic incentives to promote behavioral change. *CESifo Working Paper No. 8070*, 2020.

[57] Saeed Boorboor, Saad Nadeem, Ji Hwan Park, Kevin Baker, and Arie Kaufman. Crowdsourcing lung nodules detection and annotation. In *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, volume 10579, pages 342–348. SPIE, 2018.

[58] Walter C Borman, Rodney L Rosse, and Norman M Abrahams. An empirical construct validity approach to studying predictor–job performance links. *Journal of Applied Psychology*, 65(6):662, 1980.

[59] Ioannis Boutsis and Vana Kalogeraki. On task assignment for real-time reliable crowdsourcing. In *2014 IEEE 34th International Conference on Distributed Computing Systems*, pages 1–10. IEEE, 2014.

[60] James L Bowditch, Anthony F Buono, and Marcus M Stewart. *A primer on organizational behavior*. John Wiley & Sons, 2007.

[61] Clint A Bowers, Florian Jentsch, Eduardo Salas, and Curt C Braun. Analyzing communication sequences for team training needs assessment. *Human factors*, 40(4):672–679, 1998.

[62] Russell Boyd and Terry Brown. Pilot study of myers briggs type indicator personality profiling in emergency department senior medical staff. *Emergency Medicine Australasia*, 17(3):200–203, 2005.

[63] Gregory J Boyle. Myers-briggs type indicator (mbti): some psychometric limitations. *Australian Psychologist*, 30:71–74, 1995.

[64] Daren C Brabham. Moving the crowd at threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, 13(8): 1122–1145, 2010.

[65] Bret H Bradley, John E Baur, Christopher G Banford, and Bennett E Postlethwaite. Team players and collective performance: How agreeableness affects team performance over time. *Small Group Research*, 44(6):680–711, 2013.

[66] Steffen Brauer and Thomas C Schmidt. Group formation in elearning-enabled online social networks. In *2012 15th international conference on interactive collaborative learning (ICL)*, pages 1–8. IEEE, 2012.

[67] Virginia Braun and Victoria Clarke. *Thematic analysis*. American Psychological Association, 2012.

[68] Mark A Brennan and Courtney G Flint. Uncovering the hidden dimensions of rural disaster mitigation: Capacity building through community emergency response teams. *Journal of Rural Social Sciences*, 22(2):7, 2007.

[69] Susan M Briggs. Disaster management teams. *Current opinion in critical care*, 11 (6):585–589, 2005.

[70] Tina Budzise-Weaver, Jiangping Chen, and Mikhaela Mitchell. Collaboration and crowdsourcing: The cases of multilingual digital libraries. *The Electronic Library*, 30(2):220–232, 2012.

[71] Martin J Burnham, Yen K Le, and Ralph L Piedmont. Who is mturk? personal characteristics and sample consistency of these online workers. *Mental Health, Religion & Culture*, 21(9-10):934–944, 2018.

[72] Aleksejs Busarovs. Ethical aspects of crowdsourcing, or is it a modern form of exploitation. 2013.

[73] James N Butcher. Minnesota multiphasic personality inventory. *The Corsini Encyclopedia of Psychology*, pages 1–3, 2010.

[74] Kris Byron, Sejin Keem, Tanja Darden, Christina E Shalley, and Jing Zhou. Building blocks of idea generation and implementation in teams: A meta-analysis of team design and team creativity and innovation. *Personnel Psychology*, 76(1):249–278, 2023.

[75] Mingzhi Cai, Erin N Seligsohn, Zhinan Zhang, and Yan Wang. Risk perception and product safety in user-centered design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 59278, page V007T06A041. American Society of Mechanical Engineers, 2019.

[76] Michael A Campion, Gina J Medsker, and A Catherine Higgs. Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel psychology*, 46(4):823–847, 1993.

[77] Robert M Capraro and Mary Margaret Capraro. Myers-briggs type indicator score reliability across: Studies a meta-analytic reliability generalization study. *Educational and Psychological Measurement*, 62(4):590–602, 2002.

[78] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15, 2019.

[79] Ana Caraban, Loukas Konstantinou, and Evangelos Karapanos. The nudge deck: A design support tool for technology-mediated nudging. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pages 395–406, 2020.

[80] Mifan Careem, Chamindra De Silva, Ravindra De Silva, Louiqa Raschid, and Sanjiva Weerawarana. Sahana: Overview of a disaster management system. In *2006 International Conference on Information and Automation*, pages 361–366. IEEE, 2006.

[81] Nathan T Carter, Joshua D Miller, and Thomas A Widiger. Extreme personalities at work and in life. *Current Directions in Psychological Science*, 27(6):429–436, 2018.

[82] Liz Carver and Murray Turoff. Human-computer interaction: the human and

computer as a team in emergency management information systems. *Communications of the ACM*, 50(3):33–38, 2007.

[83] Milly Casey-Campbell and Martin L Martens. Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews*, 11(2):223–246, 2009.

[84] Leo Anthony Celi, Andrea Ippolito, Robert A Montgomery, Christopher Moses, and David J Stone. Crowdsourcing knowledge discovery and innovations in medicine. *Journal of medical Internet research*, 16(9):e3761, 2014.

[85] Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(6):1–168, 2011.

[86] Jonathan Chang, Itamar Rosenn, Lars Backstrom, and Cameron Marlow. epluribus: Ethnicity on social networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, 2010.

[87] Sofia Chatzi, Ioannis Nikolaou, and Neil Anderson. Team personality composition and team innovation implementation: The mediating role of team climate for innovation. *Applied Psychology*, 72(2):769–796, 2023.

[88] Marian M Chau. Rapid response to a tree seed conservation challenge in hawai ‘i through crowdsourcing, citizen science, and community engagement. *Journal of Sustainable Forestry*, pages 1–19, 2020.

[89] Alicia Fedelina Chavez and Florence Guido-DiBrito. Racial and ethnic identity and development. *New directions for adult and continuing education*, 84:39–44, 1999.

[90] Claire Chen. A review examining biases in workplace hiring and promotion processes. 2023.

[91] Frank X Chen, Abby C King, and Eric B Hekler. " healthifying" exergames: improving health outcomes through intentional priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1855–1864, 2014.

[92] Jilin Chen, Gary Hsieh, Jalal U Mahmud, and Jeffrey Nichols. Understanding individuals’ personal values from social media word use. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 405–414, 2014.

[93] Kailong Chen, Tianqi Chen, Guoqing Zheng, Ou Jin, Enpeng Yao, and Yong Yu. Collaborative personalized tweet recommendation. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 661–670, 2012.

[94] Liang Chen, Pei Xu, and De Liu. Effect of crowd voting on participation in crowdsourcing contests. *Journal of Management Information Systems*, 37(2): 510–535, 2020.

[95] Lichun Chen and Elise Miller-Hooks. Optimal team deployment in urban search and rescue. *Transportation Research Part B: Methodological*, 46(8):984–999, 2012.

[96] Quanze Chen, Jonathan Bragg, Lydia B Chilton, and Dan S Weld. Cicero: Multi-turn, contextual argumentation for accurate crowdsourcing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[97] Rui Chen, Raj Sharman, H Raghav Rao, and Shambhu J Upadhyaya. Coordination in emergency response management. *Communications of the ACM*, 51(5):66–73, 2008.

[98] Xiao Chen and Kaiqi Xiong. A payment scheme in crowdsourcing. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2017.

[99] Justin Cheng, Jaime Teevan, Shamsi T Iqbal, and Michael S Bernstein. Break it down: A comparison of macro-and microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 4061–4064, 2015.

[100] Yijie Cheng and Cédric Langbort. A model of informational nudging in transportation networks. In *2016 IEEE 55th conference on decision and control (CDC)*, pages 7598–7604. IEEE, 2016.

[101] Christos E Christodoulopoulos and K Papanikolaou. Investigation of group formation using low complexity algorithms. In *Proc. of PING Workshop*, pages 57–60. Citeseer, 2007.

[102] Morgane Ciot, Morgan Sonderegger, and Derek Ruths. Gender inference of twitter users in non-english contexts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, 2013.

[103] Derrick L Cogburn. Diversity matters, even at a distance: Evaluating the impact of computer-mediated communication on civil society participation in the world summit on the information society. *Information Technologies & International Development*, 1(3):pp–15, 2004.

[104] Claudia C Cogliser, William L Gardner, Mark B Gavin, and J Christian Broberg. Big five personality factors and leader emergence in virtual teams: Relationships with team trustworthiness, member performance contributions, and team performance. *Group & Organization Management*, 37(6):752–784, 2012.

[105] Raviv Cohen and Derek Ruths. Classifying political orientation on twitter: It's not easy! In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 2013.

[106] Aria X Collopy, Heather Lynn Johnson, Katherine A Goodman, Tom Altman, Maryam Darbeheshti, Kristin L Wood, and David C Mays. Exploring nudging approaches for growing a culture of diversity and inclusion with engineering faculty. In *Annual Conference, American Society for Engineering Education, Minneapolis, Minnesota, June 26-29, 2022*, 2022.

[107] Ana Colovic, Annalisa Caloffi, and Federica Rossi. Crowdsourcing and covid-19:

How public administrations mobilize crowds to find solutions to problems posed by the pandemic. *Public Administration Review*, 2022.

[108] Corey Columb and E Ashby Plant. Revisiting the obama effect: Exposure to obama reduces implicit prejudice. *Journal of Experimental Social Psychology*, 47 (2):499–501, 2011.

[109] Noshir Contractor. Who would you like to work with?: Use of individual characteristics and social networks in team formation systems. 2019.

[110] Noshir Contractor, Leslie Ann DeChurch, Anup Sawant, and Xiang Li. My dream team assembler. 2013.

[111] Alexandra Cook, Alexander Zill, and Bertolt Meyer. Perceiving leadership structures in teams: Effects of cognitive schemas and perceived communication. *Small Group Research*, page 1046496420950480, 2020.

[112] Nancy J Cooke and Margaret L Hilton. Enhancing the effectiveness of team science. 2015.

[113] Jonathan R Corney, Carmen Torres-Sánchez, A Prasanna Jagadeesan, and William C Regli. Outsourcing labour to the cloud. *International Journal of Innovation and Sustainable Development*, 4(4):294–313, 2009.

[114] Paul T Costa Jr and Robert R McCrae. *The Revised Neo Personality Inventory (neo-pi-r)*. Sage Publications, Inc, 2008.

[115] Maureen A Craig and Jennifer A Richeson. More diverse yet less tolerant? how the increasingly diverse racial landscape affects white americans' racial attitudes. *Personality and Social Psychology Bulletin*, 40(6):750–761, 2014.

[116] Richard J Crisp, Laura M Bache, and Angela T Maitner. Dynamics of social comparison in counter-stereotypic domains: Stereotype boost, not stereotype threat, for women engineering majors. *Social Influence*, 4(3):171–184, 2009.

[117] Kelly Cuccolo, Megan S Irgens, Martha S Zlokovich, Jon Grahe, and John E Edlund. What crowdsourcing can offer to cross-cultural psychological science. *Cross-Cultural Research*, 55:3–28, 2021.

[118] Jeff Cummings and Alan R Dennis. Virtual first impressions matter: the effect of enterprise social networking sites on impression formation in virtual teams. *MIS Quarterly*, 42(3):697–718, 2018.

[119] Petru Lucian Curşeu, Remus Ilies, Delia Vîrgă, Laurenţiu Maricuţoiu, and Florin A Sava. Personality characteristics that are valued in teams: Not always "more is better"? *International Journal of Psychology*, 54(5):638–649, 2019.

[120] Kristina B Dahlin, Laurie R Weingart, and Pamela J Hinds. Team diversity and information use. *Academy of management journal*, 48(6):1107–1123, 2005.

[121] Jeffrey Dalton, Victor Ajayi, and Richard Main. Vote goat: Conversational movie recommendation. In *The 41st international acm sigir conference on research & development in information retrieval*, pages 1285–1288, 2018.

[122] Nilanjana Dasgupta. Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social justice research*, 17(2):143–169, 2004.

[123] Nilanjana Dasgupta and Shaki Asgari. Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of experimental social psychology*, 40(5):642–658, 2004.

[124] Nilanjana Dasgupta and Anthony G Greenwald. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of personality and social psychology*, 81(5):800, 2001.

[125] Kemal Davaslioglu, Bob Pokorny, Yalin E Sagduyu, Henrik Molintas, Sohraab Soltani, Rebecca Grossman, and Clint Bowers. Measuring the collective allostatic load. In *2019 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, pages 93–99. IEEE, 2019.

[126] Alexander Philip Dawid and Allan M Skene. Maximum likelihood estimation of observer error-rates using the em algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28:20–28, 1979.

[127] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 2013.

[128] Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 626–638, 2014.

[129] Eider Arantes De Oliveira, Marcio Lopes Pimenta, Per Hilletofth, and David Eriksson. Integration through cross-functional teams in a service company. *European Business Review*, 2016.

[130] Triparna De Vreede, Cuong Nguyen, Gert-Jan De Vreede, Imed Boughzala, Onook Oh, and Roni Reiter-Palmon. A theoretical model of user engagement in crowdsourcing. In *Collaboration and Technology: 19th International Conference, CRIWG 2013, Wellington, New Zealand, October 30–November 1, 2013, Proceedings 19*, pages 94–109. Springer, 2013.

[131] Reinout E De Vries, Bart Van den Hooff, and Jan A de Ridder. Explaining knowledge sharing: The role of team communication styles, job satisfaction, and performance beliefs. *Communication research*, 33(2):115–135, 2006.

[132] Frank RC De Wit, Lindred L Greer, and Karen A Jehn. The paradox of intragroup conflict: a meta-analysis. *Journal of applied psychology*, 97(2):360, 2012.

[133] Fabrice Delice, Moira Rousseau, and Jennifer Feitosa. Advancing teams research: What, when, and how to measure team dynamics over time. *Frontiers in psychology*, 10:1324, 2019.

[134] Ad Den Otter and Stephen Emmitt. Exploring effectiveness of team communica-

tion: Balancing synchronous and asynchronous communication in design teams. *Engineering, Construction and Architectural Management*, 2007.

[135] Jia Deng, Jonathan Krause, and Li Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2013.

[136] Xuefei Deng, Kshiti D Joshi, and Robert D Galliers. The duality of empowerment and marginalization in microtask crowdsourcing. *MIS quarterly*, 40(2):279–302, 2016.

[137] Qiming Diao and Jing Jiang. A unified model for topics, events and users on twitter. ACL, 2013.

[138] Cristina Díaz-García, Angela González-Moreno, and Francisco Jose Saez-Martinez. Gender diversity within r&d teams: Its impact on radicalness of innovation. *Innovation*, 15(2):149–160, 2013.

[139] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 135–143, 2018.

[140] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*, pages 238–247, 2015.

[141] John M Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41:417–440, 1990.

[142] Raheleh B Dilmaghani and Ramesh R Rao. On designing communication networks for emergency situations. In *2006 IEEE International Symposium on Technology and Society*, pages 1–8. IEEE, 2006.

[143] Cheng Ding, Fan Xia, Gopakumar Gopalakrishnan, Weining Qian, and Aoying Zhou. Teamgen: an interactive team formation system based on professional social network. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 195–199, 2017.

[144] Indika Dissanayake, Sridhar P Nerur, and Jie Zhang. Team formation and performance in online crowdsourcing competitions: The role of homophily and diversity in solver characteristics. In *ICIS*, 2019.

[145] Anhai Doan, Raghu Ramakrishnan, and Alon Y Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011.

[146] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 1013–1022, 2012.

[147] Hendrik Drachsler and Paul A Kirschner. Learner characteristics. 2011.

[148] Ryan Drapeau, Lydia Chilton, Jonathan Bragg, and Daniel Weld. Microtalk: Using argumentation to improve crowdsourcing accuracy. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 4, pages 32–41, 2016.

[149] T Driskell, S Burke, JE Driskell, E Salas, and L Neuberger. Steeling the team: Assessing individual and team functioning "at a distance.". *The Military Psychologist*, 29:12–18, 2014.

[150] Tripp Driskell, Eduardo Salas, and James E Driskell. Teams in extreme environments: Alterations in team development and teamwork. *Human Resource Management Review*, 28(4):434–449, 2018.

[151] Arlene Arlene Brigoli Ducao. *OpenIR [Open Infrared]: enhancing environmental monitoring through accessible remote sensing, in Indonesia and beyond*. PhD thesis, Massachusetts Institute of Technology, 2013.

[152] P Christopher Earley. Self or group? cultural effects of training on self-efficacy and performance. *Administrative science quarterly*, pages 89–117, 1994.

[153] Dov Eden. Pygmalion without interpersonal contrast effects: Whole groups gain from raising manager expectations. *Journal of Applied Psychology*, 75(4):394, 1990.

[154] Bryan D Edwards, Eric Anthony Day, Winfred Arthur Jr, and Suzanne T Bell. Relationships among team ability composition, team mental models, and team performance. *Journal of applied psychology*, 91(3):727, 2006.

[155] Carsten Eickhoff and Arjen de Vries. How crowdsourcable is your task. In *Proceedings of the workshop on crowdsourcing for search and data mining (CSDM) at the fourth ACM international conference on web search and data mining (WSDM)*, pages 11–14, 2011.

[156] Aleksander PJ Ellis, Bradford S Bell, Robert E Ployhart, John R Hollenbeck, and Daniel R Ilgen. An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel psychology*, 58 (3):641–672, 2005.

[157] Niklas Elmqvist, Andrew Vande Moere, Hans-Christian Jetter, Daniel Cernea, Harald Reiterer, and TJ Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340, 2011.

[158] Fatma Elsafoury. Teargas, water cannons and twitter: A case study on detecting protest repression events in turkey 2013. In *Text2Story@ ECIR*, pages 5–13, 2020.

[159] Jan J Elshout and Marcel VJ Veenman. Relation between intellectual ability and working method as predictors of learning. *The Journal of Educational Research*, 85(3):134–143, 1992.

[160] Christoph Ernst, Andreas Mladenow, and Christine Strauss. Collaboration and crowdsourcing in emergency management. *International Journal of Pervasive Computing and Communications*, 2017.

[161] James C Etheridge, Rachel Moyal-Smith, Yves Sonnay, Mary E Brindle, Tze Tein Yong, Hiang Khoon Tan, Christine Lim, and Joaquim M Havens. Non-technical skills in surgery during the covid-19 pandemic: An observational study. *International Journal of Surgery*, page 106210, 2022.

[162] Charles R Evans and Kenneth L Dion. Group cohesion and performance: A meta-analysis. *Small group research*, 22(2):175–186, 1991.

[163] Johan Farkas and Christina Neumayer. ‘stop fake hate profiles on facebook’: Challenges for crowdsourced activism on social media. *First Monday*, 2017.

[164] Somayeh Fatahi, Hadi Moradi, and Leila Kashani-Vahid. A survey of personality and learning styles models applied in virtual environments with emphasis on e-learning environments. *Artificial Intelligence Review*, 46(3):413–429, 2016.

[165] Edward A Feigenbaum and Herbert A Simon. A theory of the serial position effect. *British Journal of Psychology*, 53(3):307–320, 1962.

[166] Idoya Ferrero-Ferrero, M Ángeles Fernández-Izquierdo, and M Jesús Muñoz-Torres. Age diversity: An empirical study in the board of directors. *Cybernetics and Systems*, 46(3-4):249–270, 2015.

[167] Sharon Birkman Fink and Stephanie Capparell. *The Birkman method: Your personality at work*. John Wiley & Sons, 2013.

[168] Ailbhe Finnerty, Pavel Kucherbaev, Stefano Tranquillini, and Gregorio Convertino. Keep it simple: Reward and task design in crowdsourcing. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*, pages 1–4, 2013.

[169] Annette Flanagin, Tracy Frey, Stacy L Christiansen, AMA Manual of Style Committee, et al. Updated guidance on the reporting of race and ethnicity in medical and science journals. *Jama*, 326(7):621–627, 2021.

[170] Rhona Flin and Georgina Slaven. Personality and emergency command ability. *Disaster Prevention and Management: An International Journal*, 1996.

[171] Jacqueline Floch, Michael Angermann, Edel Jennings, and Mark Roddy. Exploring cooperating smart spaces for efficient collaboration in disaster management. In *ISCRAM*, 2012.

[172] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):2, 2002.

[173] Shandra Forrest-Bank and Jeffrey M Jenson. Differences in experiences of racial and ethnic microaggression among asian, latino/hispanic, black, and white young adults. *J. Soc. & Soc. Welfare*, 42:141, 2015.

[174] Jens Förster and Nira Liberman. Knowledge activation. *Social psychology: Handbook of basic principles*, 2:201–231, 2007.

[175] World Economic Forum. The global gender gap report 2021. 2021.

[176] H Clayton Foushee. Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist*, 39(8):885, 1984.

[177] Lorie A Fridell. Racial aspects of police shootings: Reducing both bias and counter bias. *Criminology & Pub. Pol'y*, 15:481, 2016.

[178] Alexandra M Friede. In defence of the baltic sea region:(non-) allied policy responses to the exogenous shock of the ukraine crisis. *European Security*, pages 1–23, 2022.

[179] Joseph L Gabbard, Deborah Hix, and J Edward Swan. User-centered design and evaluation of virtual environments. *IEEE computer Graphics and Applications*, 19 (6):51–59, 1999.

[180] Ujwal Gadiraju, Gianluca Demartini, Ricardo Kawase, and Stefan Dietze. Human beyond the machine: Challenges and opportunities of microtask crowdsourcing. *IEEE Intelligent Systems*, 30(4):81–85, 2015.

[181] Ujwal Gadiraju, Sebastian Möller, Martin Nöllenburg, Dietmar Saupe, Sebastian Egger-Lampl, Daniel Archambault, and Brian Fisher. Crowdsourcing versus the laboratory: towards human-centered experiments using the crowd. In *Evaluation in the Crowd. Crowdsourcing and human-centered experiments*, pages 6–26. Springer, 2017.

[182] Jay R Galbraith and Edward E Lawler. *Organizing for the future: The new logic for managing complex organizations*. Jossey-Bass Inc Pub, 1993.

[183] Michael J Gallivan. Examining it professionals' adaptation to technological change: The influence of gender and personal attributes. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 35(3):28–49, 2004.

[184] Huiji Gao, Jalal Mahmud, Jilin Chen, Jeffrey Nichols, and Michelle Zhou. Modeling user attitude toward controversial topics in online social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 2014.

[185] Heidi K Gardner. Performance pressure as a double-edged sword: Enhancing team motivation but undermining the use of team knowledge. *Administrative Science Quarterly*, 57:1–46, 2012.

[186] Bertram Gawronski, Roland Deutsch, Sawsan Mbirkou, Beate Seibt, and Fritz Strack. When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of experimental social psychology*, 44(2):370–377, 2008.

[187] Miriam Gay-Antaki. Stories from the ipcc: An essay on climate science in fourteen questions. *Global Environmental Change*, 71:102384, 2021.

[188] David Geiger, Michael Rosemann, and Erwin Fielt. Crowdsourcing information systems–a systems theory perspective. 2011.

[189] David Geiger, Stefan Seedorf, Thimo Schulze, Robert C Nickerson, and Martin

Schader. Managing the crowd: Towards a taxonomy of crowdsourcing processes. In *AMCIS*, 2011.

[190] Leslie H Gelling. Applying for ethical approval for research: the main issues. *Nursing Standard*, 30(20):40–44, 2016.

[191] Cristina Gena, Pierluigi Grillo, Antonio Lieto, Claudio Mattutino, and Fabiana Vernero. When personalization is not an option: an in-the-wild study on persuasive news recommendation. *Information*, 10(10):300, 2019.

[192] Connie JG Gersick and J Richard Hackman. Habitual routines in task-performing groups. *Organizational behavior and human decision processes*, 47:65–97, 1990.

[193] Josette MP Gevers and Miranda AG Peeters. A pleasure working together? the effects of dissimilarity in team member conscientiousness on team temporal processes and individual satisfaction. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 30(3):379–400, 2009.

[194] Josette MP Gevers and Miranda AG Peeters. A pleasure working together the effects of dissimilarity in team member conscientiousness on team temporal processes and individual satisfaction. 30(3):379–400, 2009.

[195] Carol Gill, Isabel Metz, Amanuel G Tekleab, and Ian O Williamson. The combined role of conscientiousness, social networks, and gender diversity in explaining individual performance in self-managed teams. *Journal of Business Research*, 106: 250–260, 2020.

[196] Jerry W Gilley, M Lane Morris, Alina M Waite, Tabitha Coates, and Abigail Veliquette. Integrated theoretical model for building effective teams. *Advances in Developing Human Resources*, 12:7–28, 2010.

[197] Peter A Gloor. *Swarm creativity: Competitive advantage through collaborative innovation networks.* Oxford University Press, 2006.

[198] Małgorzata A Gocłowska, Richard J Crisp, and Kirsty Labuschagne. Can counter-stereotypes boost flexible thinking? *Group Processes & Intergroup Relations*, 16(2): 217–231, 2013.

[199] Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems*, pages 253–262. 2011.

[200] Lewis R Goldberg. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.

[201] Daniel G Goldstein, Eric J Johnson, Andreas Herrmann, and Mark Heitmann. Nudge your customers toward better choices. *Harvard Business Review*, 86(12): 99–105, 2008.

[202] Diego Gómez-Zará, Mengzi Guo, Leslie A DeChurch, and Noshir Contractor. The impact of displaying diversity information on the formation of self-assembling

teams. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.

[203] Wei Gong. Profiling social media users with selective self-disclosure behavior. 2016.

[204] Benjamin M Good and Andrew I Su. Crowdsourcing for bioinformatics. *Bioinformatics*, 29(16):1925–1933, 2013.

[205] Jessica J Good, Julie A Woodzicka, and Lylan C Wingfield. The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *The Journal of social psychology*, 150(2):132–147, 2010.

[206] Joseph K Goodman and Gabriele Paolacci. Crowdsourcing consumer research. *Journal of Consumer Research*, 44:196–210, 2017.

[207] C Grams. Why the open source way trumps the crowdsourcing way. *Opensource. com, available at: http://opensource. com/business/10/4/why-open-source-waytrumpscrowdsourcing-way*, 2010.

[208] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, 78 (6):1360–1380, 1973.

[209] Mary L Gray, Siddharth Suri, Syed Shoaib Ali, and Deepti Kulkarni. The crowd is a collaborative network. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pages 134–147, 2016.

[210] Daryl D Green, Jack McCann, Thao Vu, Nancy Lopez, and Samuel Ouattara. Gig economy and the future of work: A fiverr. com case study. *Management and Economics Research Journal*, 4(2):281, 2018.

[211] Paul Green and Lisa Wei-Haas. The rapid development of user interfaces: Experience with the wizard of oz method. In *Proceedings of the Human Factors Society Annual Meeting*, volume 29, pages 470–474. SAGE Publications Sage CA: Los Angeles, CA, 1985.

[212] Robin Guers, Cedric Langbort, and Dan Work. On informational nudging and control of payoff-based learning. *IFAC Proceedings Volumes*, 46(27):69–74, 2013.

[213] Yves RF Guillaume, Jeremy F Dawson, Lilian Otaye-Ebede, Stephen A Woods, and Michael A West. Harnessing demographic differences in organizations: What moderates the effects of workplace diversity? *Journal of Organizational Behavior*, 38(2):276–303, 2017.

[214] Arpit Gupta. Accelerating remote work after covid-19. *The Center for Growth and Opportunity*, 2022.

[215] Richard A Guzzo and Marcus W Dickson. Teams in organizations: Recent research on performance and effectiveness. *Annual review of psychology*, 47:307–338, 1996.

[216] Richard A Guzzo and Gregory P Shea. Group performance and intergroup relations in organizations. 1992.

[217] J Richard Hackman. *Groups that work and those that don't.* Number E10 H123. Jossey-Bass, 1990.

[218] J Richard Hackman and Charles G Morris. Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In *Advances in experimental social psychology*, volume 8, pages 45–99. Elsevier, 1975.

[219] Matthew W Hahn and R Alexander Bentley. Drift as a mechanism for cultural change: an example from baby names. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(suppl_1):S120–S123, 2003.

[220] Jerayr Haleblian and Sydney Finkelstein. Top management team size, ceo dominance, and firm performance: The moderating roles of environmental turbulence and discretion. *Academy of management journal*, 36(4):844–863, 1993.

[221] Lei Han, Kevin Roitero, Ujwal Gadiraju, Cristina Sarasua, Alessandro Checco, Eddy Maddalena, and Gianluca Demartini. All those wasted hours: On task abandonment in crowdsourcing. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 321–329, 2019.

[222] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 1914–1933, 2017.

[223] Derek L Hansen, Patrick J Schone, Douglas Corey, Matthew Reid, and Jake Gehring. Quality control mechanisms for crowdsourcing: peer review, arbitration, & expertise at familysearch indexing. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 649–660, 2013.

[224] Karin Hansson and Thomas Ludwig. Crowd dynamics: Conflicts, contradictions, and community in crowdsourcing. *Computer Supported Cooperative Work (CSCW)*, 28:791–794, 2019.

[225] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[226] Albert A Harrison and Mary M Connors. Groups in exotic environments. In *Advances in experimental social psychology*, volume 18, pages 49–87. Elsevier, 1984.

[227] David A Harrison and Katherine J Klein. What's the difference? diversity constructs as separation, variety, or disparity in organizations. *Academy of management review*, 32(4):1199–1228, 2007.

[228] David A Harrison, Kenneth H Price, and Myrtle P Bell. Beyond relational demography: Time and the effects of surface-and deep-level diversity on work group cohesion. *Academy of management journal*, 41:96–107, 1998.

[229] Jonathon D Hart, Thammathip Piumsomboon, Louise Lawrence, Gun A Lee, Ross T Smith, and Mark Billinghurst. 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*, pages 117–120, 2018.

[230] Reid Hastie, Colin Schroeder, and Renée Weber. Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society*, 28(3): 242–247, 1990.

[231] Christian Heath and Paul Luff. Collaboration and controlcrisis management and multimedia technology in london underground line control rooms. *Computer Supported Cooperative Work (CSCW)*, 1:69–94, 1992.

[232] Robert L Heath and Michael Palenchar. Community relations and risk communication: A longitudinal study of the impact of emergency response messages. *Journal of Public Relations Research*, 12(2):131–161, 2000.

[233] Jessica Heinzelman and Carol Waters. *Crowdsourcing crisis information in disaster-affected Haiti*. US Institute of Peace Washington, DC, 2010.

[234] Natali Helberger. Diversity by design. *Journal of Information Policy*, 1:441–469, 2011.

[235] Robert L Helmreich. Prolonged stress in sealab ii: A field study of individual and group reactions. Technical report, YALE UNIV NEW HAVEN CT, 1967.

[236] Libby Hemphill and Andrew J Roback. Tweet acts: how constituents lobby congress via twitter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work and social computing*, pages 1200–1210, 2014.

[237] Rebecca Henderson. *Reimagining Capitalism in a World on Fire: Shortlisted for the FT & McKinsey Business Book of the Year Award 2020*. Penguin UK, 2021.

[238] Andreas Herrmann, Daniel G Goldstein, Rupert Stadler, Jan R Landwehr, Mark Heitmann, Reto Hofstetter, and Frank Huber. The effect of default options on choice—evidence from online product configurators. *Journal of Retailing and Consumer Services*, 18(6):483–491, 2011.

[239] Jocelyn J. Herstein, Michelle M. Schwedhelm, Angela Vasa, Paul D. Biddinger, and Angela L. Hewlett. Emergency preparedness: What is the future? *Antimicrobial Stewardship Healthcare Epidemiology*, 1:e29, 2021. doi: 10.1017/ash.2021.190.

[240] Lars Hetmank. Components and functions of crowdsourcing systems–a systematic literature review. 2013.

[241] Danula Hettiachchi, Vassilis Kostakos, and Jorge Goncalves. A survey on task assignment in crowdsourcing. *ACM Computing Surveys (CSUR)*, 55(3):1–35, 2022.

[242] Scott Highhouse, Yi Wang, and Don C Zhang. Is risk propensity unique from the big five factors of personality? a meta-analytic investigation. *Journal of Research in Personality*, page 104206, 2022.

[243] Pamela J Hinds and Catherine Durnell Cramton. Situated coworker familiarity: How site visits transform relationships among distributed workers. *Organization Science*, 25(3):794–814, 2014.

[244] Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*, pages 419–429, 2015.

[245] Joses Ho, Tayfun Tumkaya, Sameer Aryal, Hyungwon Choi, and Adam Claridge-Chang. Moving beyond p values: data analysis with estimation graphics. *Nature methods*, 16(7):565–566, 2019.

[246] Christian Hugo Hoffmann. Is ai intelligent? an assessment of artificial intelligence, 70 years after turing. *Technology in Society*, 68:101893, 2022.

[247] John R Hollenbeck, Bianca Beersma, and Maartje E Schouten. Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Academy of Management Review*, 37(1):82–106, 2012.

[248] Jessica Höpfner and Nina Keith. Goal missed, self hit: Goal-setting, goal-failure, and their affective, motivational, and behavioral consequences. *Frontiers in psychology*, 12:704790, 2021.

[249] Akbar Hossain, Farhaan Mirza, M Asif Naeem, and Jairo Gutierrez. A crowd sourced framework for neighbour assisted medical emergency system. In *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)*, pages 1–6. IEEE, 2017.

[250] Jeff Howe et al. The rise of crowdsourcing. *Wired magazine*, 14(6):176–183, 2006.

[251] Shih-Wen Huang and Wai-Tat Fu. Don't hide in the crowd! increasing social transparency between peer workers improves crowdsourcing outcomes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 621–630, 2013.

[252] John A Hughes, David Randall, and Dan Shapiro. Faltering from ethnography to design. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 115–122, 1992.

[253] Dennis Hummel and Alexander Maedche. How effective is nudging? a quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80:47–58, 2019.

[254] Vivian Hunt, Dennis Layton, and Sara Prince. Diversity matters. *McKinsey & Company*, 1:15–29, 2015.

[255] Russell RC Hutter and Richard J Crisp. The composition of category conjunctions. *Personality and Social Psychology Bulletin*, 31(5):647–657, 2005.

[256] Henry Oluchukwu Ikediego, Mustafa Ilkan, A Mohammed Abubakar, and Festus Victor Bekun. Crowd-sourcing (who, why and what). *International Journal of Crowd Science*, 2:27–41, 2018.

[257] Gözde Ikizer, Marta Kowal, İlknur Dilekler Aldemir, Alma Jeftić, Aybegum Memisoglu-Sanli, Arooj Najmussaqib, David Lacko, Kristina Eichel, Fidan Turk, Stavroula Chrona, et al. Big five traits predict stress and loneliness during the covid-19 pandemic: Evidence for the role of neuroticism. *Personality and Individual Differences*, page 111531, 2022.

[258] Indeed. 25 most common jobs in america, Nov 2020. URL `https://www.indeed.com/career-advice/finding-a-job/most-common-jobs-in-america`.

[259] Panagiotis G Ipeirotis. Mechanical turk, low wages, and the market for lemons. *A Computer Scientist in a Business School*, 27, 2010.

[260] Robert W Irving. An efficient algorithm for the "stable roommates" problem. *Journal of Algorithms*, 6(4):577–595, 1985.

[261] Susan E Jackson. The consequences of diversity in multidisciplinary work teams. *Handbook of work group psychology*, pages 53–75, 1996.

[262] David Jacobs and Larry Singell. Leadership and organizational performance: Isolating links between managers and collective success. *Social Science Research*, 22(2):165–189, 1993.

[263] A Prasanna Jagadeesan, Andrew Lynn, Jonathan R Corney, XT Yan, Jan Wenzel, Andrew Sherlock, and W Regli. Geometric reasoning via internet crowdsourcing. In *2009 SIAM/ACM Joint Conference on Geometric and Physical Modeling*, pages 313–318, 2009.

[264] Farnaz Jahanbakhsh, Wai-Tat Fu, Karrie Karahalios, Darko Marinov, and Brian Bailey. You want me to work with who? stakeholder perceptions of automated team formation in project-based courses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3201–3212, 2017.

[265] Niklas Jakobsson and Henrik Lindholm. Ethnic preferences in internet dating: A field experiment. *Marriage & Family Review*, 50(4):307–317, 2014.

[266] Wiebren S Jansen, Sabine Otten, and Karen I van der Zee. Being part of diversity: The effects of an all-inclusive multicultural diversity approach on majority members' perceived inclusion and support for organizational diversity efforts. *Group Processes & Intergroup Relations*, 18(6):817–832, 2015.

[267] Sabeen Javaid, Awais Majeed, and Hammad Afzal. A reputation management system for efficient selection of disaster management team. In *2013 15th International Conference on Advanced Communications Technology (ICACT)*, pages 829–834. IEEE, 2013.

[268] Debora Jeske and Kenneth S Shultz. Using social media content for screening in recruitment and selection: pros and cons. *Work, employment and society*, 30(3):535–546, 2016.

[269] Mathias Jesse and Dietmar Jannach. Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, 3:100052, 2021.

[270] Birgit Helene Jevnaker and Johan Olaisen. The future of knowledge work: Working smarter and greener in the age of digitalization. In *Reimagining Sustainable Organization*, pages 137–156. Springer, 2022.

[271] Yanli Jia, Yi Zhao, and Yuli Lin. Effects of system characteristics on users' self-disclosure in social networking sites. In *2010 Seventh International Conference on Information Technology: New Generations*, pages 529–533. IEEE, 2010.

[272] Yuchao Jiang, Daniel Schlagwein, and Boualem Benatallah. A review on crowdsourcing for education: State of the art of literature and practice. *PACIS*, page 180, 2018.

[273] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. Beyond nudges: Tools of a choice architecture. *Marketing letters*, 23:487–504, 2012.

[274] Heather A Johnson. Slack. *Journal of the Medical Library Association: JMLA*, 106: 148, 2018.

[275] Martin R Kalfatovic, Effie Kapsalis, Katherine P Spiess, Anne Van Camp, and Michael Edson. Smithsonian team flickr: a library, archives, and museums collaboration in web 2.0 space. *Archival science*, 8(4):267–277, 2008.

[276] Wim Kamphuis, Anthony WK Gaillard, and Ad LW Vogelaar. The effects of physical threat on team processes during complex task performance. *Small Group Research*, 42(6):700–729, 2011.

[277] Thivya Kandappu, Arik Friedman, Vijay Sivaraman, and Roksana Boreli. Privacy in crowdsourced platforms. *Privacy in a Digital, Networked World: Technologies, Implications and Solutions*, pages 57–84, 2015.

[278] Barbara G Kanki, Sandra Lozito, and H Clayton Foushee. Communication indices of crew coordination. *Aviation, Space, and Environmental Medicine*, 1989.

[279] Barbara G Kanki, Valerie Greaud Folk, and Cheryl M Irwin. Communication variations and aircrew performance. *The International Journal of Aviation Psychology*, 1(2):149–162, 1991.

[280] Rea Karachiwalla and Felix Pinkow. Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. *Creativity and innovation management*, 30(3):563–584, 2021.

[281] Evgeny Karataev and Vladimir Zadorozhny. Adaptive social learning based on crowdsourcing. *IEEE Transactions on Learning Technologies*, 10(2):128–139, 2016.

[282] Randi Karlsen and Anders Andersen. Recommendations with a nudge. *Technologies*, 7(2):45, 2019.

[283] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

[284] Kristiina Karvonen[1], Sanna Shibasaki[1], Sofia Nunes[1], Puneet Kaur[1], and Olli Immonen. Visual nudges for enhancing the use and produce of reputation information. 2010.

[285] Shinji Kawatsuma, Mineo Fukushima, and Takashi Okada. Emergency response by robots to fukushima-daiichi accident: summary and lessons learned. *Industrial Robot: An International Journal*, 2012.

[286] Alan E Kazdin and Richard R Bootzin. The token economy: An evaluative review 1. *Journal of applied behavior analysis*, 5(3):343–372, 1972.

[287] Belinda Kennedy, Kate Curtis, and Donna Waters. The personality of emergency nurses: is it unique? *Australasian Emergency Nursing Journal*, 17(4):139–145, 2014.

[288] Imran Khan, Ismail Khan, and Ismail Senturk. Board diversity and quality of csr disclosure: evidence from pakistan. *Corporate Governance: The International Journal of Business in Society*, 2019.

[289] Vassilis-Javed Khan, Ioanna Lykourentzou, and Georgios Metaxas. A six-month, multi-platform investigation of creative crowdsourcing. In *Advances in Longitudinal HCI Research*, pages 227–242. Springer, 2021.

[290] Firas Khatib, Frank DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, Seth Cooper, Maciej Kazmierczyk, Miroslaw Gilski, Szymon Krzywda, Helena Zabranska, Iva Pichova, et al. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature structural & molecular biology*, 18 (10):1175–1177, 2011.

[291] Ali Hosseini Khayat. *Distributed Wizard of Oz Usability Testing for Agile Teams*. UNIVERSITY OF CALGARY, 2009.

[292] Susan L Kichuk and Willi H Wiesner. The big five personality factors and team performance: implications for selecting successful product design teams. *Journal of Engineering and Technology management*, 14(3-4):195–221, 1997.

[293] David Kieras, David Meyer, and James Ballas. Towards demystification of direct manipulation: Cognitive modeling charts the gulf of execution. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 128–135, 2001.

[294] Andrea Kiesel, Wilfried Kunde, and Joachim Hoffmann. Mechanisms of subliminal response priming. *Advances in Cognitive Psychology*, 3(1-2):307, 2007.

[295] Emily Kilner and Lorraine A Sheppard. The role of teamwork and communication in the emergency department: a systematic review. *International emergency nursing*, 18(3):127–137, 2010.

[296] Haewon Kim, Seung-June Hwang, and Woojin Yoon. Industry cluster, organizational diversity, and innovation. *International Journal of Innovation Studies*, 7(3): 187–195, 2023.

[297] J Peter Kincaid, Joseph Donovan, and Beth Pettitt. Simulation techniques for training emergency response. *International Journal of Emergency Management*, 1 (3):238–246, 2003.

[298] Granville King. Crisis management & team effectiveness: A closer examination. *Journal of Business Ethics*, 41(3):235–249, 2002.

[299] Frank Kleemann, G Günter Voß, and Kerstin Rieder. Un (der) paid innovators: The commercial utilization of consumer work through crowdsourcing. *Science, technology & innovation studies*, 4:5–26, 2008.

[300] Akiva Kleinerman, Ariel Rosenfeld, and Sarit Kraus. Providing explanations for recommendations in reciprocal environments. In *Proceedings of the 12th ACM conference on recommender systems*, pages 22–30, 2018.

[301] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4):441–504, 2012.

[302] Donald Knuth. Steelcrategames, 2021. URL `https://twitter.com/steelcrategames`.

[303] Ari Kobren, Chun How Tan, Panagiotis Ipeirotis, and Evgeniy Gabrilovich. Getting more for less: Optimized crowdsourcing with dynamic tasks and goals. In *Proceedings of the 24th international conference on world wide web*, pages 592–602, 2015.

[304] Alison M Konrad, Pushkala Prasad, and Judith Pringle. *Handbook of workplace diversity*. Sage, 2005.

[305] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, 110(15):5802–5805, 2013.

[306] Ziyi Kou, Lanyu Shang, Yang Zhang, and Dong Wang. Hc-covid: A hierarchical crowdsource knowledge graph approach to explainable covid-19 misinformation detection. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP): 1–25, 2022.

[307] Steve WJ Kozlowski and Daniel R Ilgen. Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7(3):77–124, 2006.

[308] Steve WJ Kozlowski and Katherine J Klein. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. 2000.

[309] Mirjam E Kretzschmar, Ben Ashby, Elizabeth Fearon, Christopher E Overton, Jasmina Panovska-Griffiths, Lorenzo Pellis, Matthew Quaife, Ganna Rozhnova, Francesca Scarabel, Helena B Stage, et al. Challenges for modelling interventions for future pandemics. *Epidemics*, page 100546, 2022.

[310] Stefan Krumm, Jens Kanthak, Kai Hartmann, and Guido Hertel. What does it take to be a virtual team player? the knowledge, skills, abilities, and other

characteristics required in virtual teams. *Human Performance*, 29(2):123–142, 2016.

[311] Nathan R Kuncel, Sarah A Hezlett, and Deniz S Ones. Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of personality and social psychology*, 86:148, 2004.

[312] Ziva Kunda, Dale T Miller, and Theresa Claire. Combining social concepts: The role of causal reasoning. *Cognitive Science*, 14(4):551–577, 1990.

[313] Craig E Kuziemsky, Elizabeth M Borycki, Mary Ellen Purkis, Fraser Black, Michael Boyle, Denise Cloutier-Fisher, Lee Ann Fox, Patricia MacKenzie, Ann Syme, Coby Tschanz, et al. An interdisciplinary team communication framework and its application to healthcare'e-teams' systems design. *BMC Medical Informatics and Decision Making*, 9:1–15, 2009.

[314] Sang Gyu Kwak and Jong Hae Kim. Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology*, 70(2):144–156, 2017.

[315] Roy B Lacoursiere. *The life cycle of groups: Group developmental stage theory*. Human Sciences Press New York, 1980.

[316] Karim R Lakhani, Anne-Laure Fayard, Natalia Levina, and Stephanie Healy Pokrywa. Openideo. *Harvard Business School Technology & Operations Mgt. Unit Case*, (612-066), 2012.

[317] Cliff Lampe, Paul Zube, Jusil Lee, Chul Hyun Park, and Erik Johnston. Crowd-sourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly*, 31(2):317–326, 2014.

[318] Jonas Landgren and Urban Nulden. A study of emergency response work: patterns of mobile phone interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1323–1332, 2007.

[319] Bruce Lansky. *25,000+ Baby Names*. Hachette UK, 2011.

[320] Bahareh Lashkari, Javad Rezazadeh, Reza Farahbakhsh, and Kumbesan Sandrasegaran. Crowdsourcing and sensing for indoor localization in iot: A review. *IEEE Sensors Journal*, 19(7):2408–2434, 2018.

[321] Gary P Latham. The effect of priming goals on organizational-related behavior. *The psychology of thinking about the future*, 392, 2018.

[322] Paul J Lavrakas. *Encyclopedia of survey research methods*. Sage Publications, 2008.

[323] Richard A Layton, Misty L Loughry, Matthew W Ohland, and George D Ricco. Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. *Advances in Engineering Education*, 2:n1, 2010.

[324] Linda Searle Leach and Ann M Mayo. Rapid response teams: qualitative analysis of their effectiveness. *American Journal of Critical Care*, 22(3):198–210, 2013.

[325] Melissa Leach, Hayley MacGregor, Santiago Ripoll, Ian Scoones, and Annie Wilkinson. Rethinking disease preparedness: incertitude and the politics of knowledge. *Critical Public Health*, 32:82–96, 2022.

[326] Daniel H Lee and Tzyy-Ping Jung. A virtual reality game as a tool to assess physiological correlations of stress. *arXiv preprint arXiv:2009.14421*, 2020.

[327] Mei-chun Lee. Free the data from the birdcage: Opening up data and crowdsourcing activism in taiwan. *PoLAR: Political and Legal Anthropology Review*, 2020.

[328] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. Mining behavioral economics to design persuasive technology for healthy choices. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 325–334, 2011.

[329] Tiane L Lee, Leigh S Wilton, and Virginia SY Kwan. Essentializing ethnicity: Identification constraint reduces diversity interest. *Journal of Experimental Social Psychology*, 55:194–200, 2014.

[330] Jan Marco Leimeister, Michael Huber, Ulrich Bretschneider, and Helmut Krcmar. Leveraging crowdsourcing: activation-supporting components for it-based ideas competition. *Journal of management information systems*, 26:197–224, 2009.

[331] Jan Marco Leimeister, Shkodran Zogaj, and David Durward. New forms of employment and it–crowdsourcing. In *4th conference of the regulating for decent work network*, 2015.

[332] Michael Leonard, Suzanne Graham, and Doug Bonacum. The human factor: the critical importance of effective teamwork and communication in providing safe care. *BMJ Quality & Safety*, 13(suppl 1):i85–i90, 2004.

[333] Thomas C Leonard. Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness: Yale university press, new haven, ct, 2008, 293 pp, 26.00, 2008.

[334] Gabriel Shing-Koon Leung, Vincent Cho, and CH Wu. Crowd workers' continued participation intention in crowdsourcing platforms: An empirical study in compensation-based micro-task crowdsourcing. *Journal of Global Information Management (JGIM)*, 29(6):1–28, 2021.

[335] Weiwen Leung, Zheng Zhang, Daviti Jibuti, Jinhao Zhao, Maximilian Klein, Casey Pierce, Lionel Robert, and Haiyi Zhu. Race, gender and beauty: The effect of information provision on online hiring biases. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2020.

[336] Sheen S Levine and Michael J Prietula. Open collaboration for innovation: Principles and performance. *Organization Science*, 25(5):1414–1433, 2014.

[337] Rui Li, Shengjie Wang, Hongbo Deng, Rui Wang, and Kevin Chen-Chuan Chang. Towards social user profiling: unified and discriminative influence model for inferring home locations. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1023–1031, 2012.

[338] Xutao Li, Gao Cong, Xiao-Li Li, Tuan-Anh Nguyen Pham, and Shonali Krishnaswamy. Rank-geofm: A ranking based geographical factorization method for point of interest recommendation. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 433–442, 2015.

[339] Xutao Li, Tuan-Anh Nguyen Pham, Gao Cong, Quan Yuan, Xiao-Li Li, and Shonali Krishnaswamy. Where you instagram? associating your instagram photos with points of interest. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1231–1240, 2015.

[340] Lizi Liao, Jing Jiang, Ying Ding, Heyan Huang, and Ee-Peng Lim. Lifetime lexical variation in social media. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

[341] Jack Lindamood, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. Inferring private information using social network data. In *Proceedings of the 18th international conference on World wide web*, pages 1145–1146, 2009.

[342] An Liu, Weiqi Wang, Shuo Shang, Qing Li, and Xiangliang Zhang. Efficient task assignment in spatial crowdsourcing with worker and task privacy protection. *GeoInformatica*, 22(2):335–362, 2018.

[343] Angli Liu, Stephen Soderland, Jonathan Bragg, Christopher H Lin, Xiao Ling, and Daniel S Weld. Effective crowd annotation for relation extraction. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 897–906, 2016.

[344] Wendy Liu and Derek Ruths. What's in a name? using first names as features for gender inference in twitter. In *2013 AAAI Spring Symposium Series*. Citeseer, 2013.

[345] Yuping Liu, Qi Liu, Runze Wu, Enhong Chen, Yu Su, Zhigang Chen, and Guoping Hu. Collaborative learning team formation: a cognitive modeling perspective. In *International Conference on Database Systems for Advanced Applications*, pages 383–400. Springer, 2016.

[346] Edwin A Locke and Gary P Latham. *A theory of goal setting & task performance.* Prentice-Hall, Inc, 1990.

[347] George Loewenstein and Nick Chater. Putting nudges in perspective. *Behavioural Public Policy*, 1(1):26–53, 2017.

[348] Patricia H Longstaff and Sung-Un Yang. Communication management and trust: their role in building resilience to "surprises" such as natural disasters, pandemic flu, and terrorism. *Ecology and Society*, 13, 2008.

[349] Linda Lopez and Adrian D Pantoja. Beyond black and white: General support for race-conscious policies among african americans, latinos, asian americans and whites. *Political Research Quarterly*, 57(4):633–642, 2004.

[350] Lin Lu, Fuli Li, Kwok Leung, Krishna Savani, and Michael W Morris. When

can culturally diverse teams be more creative? the role of leaders' benevolent paternalism. *Journal of Organizational Behavior*, 39(4):402–415, 2018.

[351] Ioanna Lykourentzou, Dimitrios J Vergados, Katerina Papadaki, and Yannick Naudet. Guided crowdsourcing for collective work coordination in corporate environments. In *International Conference on Computational Collective Intelligence*, pages 90–99. Springer, 2013.

[352] Ioanna Lykourentzou, Angeliki Antoniou, and Yannick Naudet. Matching or crashing? personality-based team formation in crowdsourcing environments. *arXiv preprint arXiv:1501.06313*, 2015.

[353] Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P. Dow. 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*, page 260–273, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450335928.

[354] Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P Dow. 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 & Social Computing*, pages 260–273, 2016.

[355] Ioanna Lykourentzou, Shannon Wang, Robert E Kraut, and Steven P Dow. Team dating: A self-organized team formation strategy for collaborative crowdsourcing. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pages 1243–1249, 2016.

[356] Ioanna Lykourentzou, Vassillis-Javed Khan, Konstantinos Papangelis, and Panos Markopoulos. Macrotask crowdsourcing: An integrated definition, 2019.

[357] Ioanna Lykourentzou, Lionel P Robert Jr, and Pierre-Jean Barlatier. Unleashing the potential of crowd work: The need for a post-taylorism crowdsourcing model. *M@ n@ gement*, 24(4):64–69, 2021.

[358] Ioanna Lykourentzou, Federica Lucia Vinella, Faez Ahmed, Costas Papastathis, Konstantinos Papangelis, Vassilis-Javed Khan, and Judith Masthoff. Self-organizing teams in online work settings. *arXiv preprint arXiv:2102.07421*, 2021.

[359] Aaron R Lyon and Kelly Koerner. User-centered design for psychosocial intervention development and implementation. *Clinical Psychology: Science and Practice*, 23(2):180, 2016.

[360] Jie Ma, Yisheng Peng, and Bo Wu. Challenging or hindering? the roles of goal orientation and cognitive appraisal in stressor-performance relationships. *Journal of Organizational Behavior*, 42(3):388–406, 2021.

[361] Nils Machts, Johanna Kaiser, Fabian TC Schmidt, and Jens Moeller. Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, 19:85–103, 2016.

[362] Barati Jozan M Mahdi and Taghiyareh Fattaneh. A semi-pareto optimal set based algorithm for grouping of students. In *4th International Conference on e-Learning and e-Teaching (ICELET 2013)*, pages 10–13. IEEE, 2013.

[363] Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Where is this tweet from? inferring home locations of twitter users. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 6, 2012.

[364] ANN Majchrzak and Philip HB More. Emergency! web 2.0 to the rescue! *Communications of the ACM*, 54(4):125–132, 2011.

[365] Ann Majchrzak, Arvind Malhotra, Ann Majchrzak, and Arvind Malhotra. What is crowdsourcing for innovation? *Unleashing the crowd: collaborative solutions to wicked business and societal problems*, pages 3–46, 2020.

[366] Sakhavat Mammadov. Big five personality traits and academic performance: A meta-analysis. *Journal of personality*, 90(2):222–255, 2022.

[367] Elizabeth Mannix and Margaret A Neale. What differences make a difference? the promise and reality of diverse teams in organizations. *Psychological science in the public interest*, 6(2):31–55, 2005.

[368] Ji-Ye Mao, Karel Vredenburg, Paul W Smith, and Tom Carey. The state of user-centered design practice. *Communications of the ACM*, 48(3):105–109, 2005.

[369] Naseebah Maqtary, Abdulqader Mohsen, and Kamal Bechkoum. Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning*, 24(2):169–190, 2019.

[370] Anthony J Marcel. Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive psychology*, 15(2):197–237, 1983.

[371] Matthew Marge, Claire Bonial, Brendan Byrne, Taylor Cassidy, A William Evans, Susan G Hill, and Clare Voss. Applying the wizard-of-oz technique to multimodal human-robot dialogue. *arXiv preprint arXiv:1703.03714*, 2017.

[372] Gloria Mark, Andrew L Kun, Sean Rintel, and Abigail Sellen. Introduction to this special issue: the future of remote work: responses to the pandemic. *Human–Computer Interaction*, pages 1–7, 2022.

[373] Shannon L Marlow, Christina N Lacerenza, Jensine Paoletti, C Shawn Burke, and Eduardo Salas. Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational behavior and human decision processes*, 144:145–170, 2018.

[374] Catherine C Marshall and Frank M Shipman. Experiences surveying the crowd: Reflections on methods, participation, and reliability. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 234–243, 2013.

[375] C Martella, J Li, C Conrado, and A Vermeeren. On current crowd management practices and the need for increased situation awareness, prediction, and intervention. *Safety science*, 91:381–393, 2017.

[376] Gonzalo Martínez, Lauren Watson, Pedro Reviriego, José Alberto Hernández, Marc Juarez, and Rik Sarkar. Combining generative artificial intelligence (ai) and the internet: Heading towards evolution or degradation? *arXiv preprint arXiv:2303.01255*, 2023.

[377] Luis L Martins and Christina E Shalley. Creativity in virtual work: Effects of demographic differences. *Small group research*, 42(5):536–561, 2011.

[378] Elaine Massung, David Coyle, Kirsten F Cater, Marc Jay, and Chris Preist. Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on human factors in Computing systems*, pages 371–380, 2013.

[379] Judith Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. In *Personalized digital television*, pages 93–141. Springer, 2004.

[380] Judith Masthoff. The user as wizard: A method for early involvement in the design and evaluation of adaptive systems. In *Fifth workshop on user-centred design and evaluation of adaptive systems*, volume 1, pages 460–469. Citeseer, 2006.

[381] David Matsumoto and Linda Juang. *Culture and psychology*. Nelson Education, 2016.

[382] Robert R McCrae. Creativity, divergent thinking, and openness to experience. *Journal of personality and social psychology*, 52(6):1258, 1987.

[383] Robert R McCrae and Paul T Costa Jr. Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, 57:17–40, 1989.

[384] Robert R McCrae and Oliver P John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.

[385] Earl H Mckinney Jr, James R Barker, Kevin J Davis, and Daryl Smith. How swift starting action teams get off the ground: What united flight 232 and airline flight crews can tell us about team communication. *Management Communication Quarterly*, 19(2):198–237, 2005.

[386] Poppy Lauretta McLeod, Jeffrey K Liker, and Sharon A Lobel. Process feedback in task groups: An application of goal setting. *The Journal of applied behavioral science*, 28:15–41, 1992.

[387] IC McManus, A Keeling, and E Paice. Stress, burnout and doctors' attitudes to work are determined by personality and learning style: a twelve year longitudinal study of uk medical graduates. *BMC medicine*, 2:29, 2004.

[388] Megan McMichael, Michael Beverly, James Noon, Tabitha Patterson, and Gary R Webb. Role improvising under conditions of uncertainty: a classification of types. 1999.

[389] Nathan McNeese, Mustafa Demir, Erin Chiou, Nancy Cooke, and Giovanni Yanikian. Understanding the role of trust in human-autonomy teaming. In *Proceedings of the 52nd Hawaii international conference on system sciences*, 2019.

[390] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27:415–444, 2001.

[391] Don Meagher. Introduction to the miller analogies test, 2006.

[392] David Mendonça. Decision support for improvisation in response to extreme events: Learning from the response to the 2001 world trade center attack. *Decision Support Systems*, 43(3):952–967, 2007.

[393] Christian Meske and Tobias Potthoff. The dinu-model–a process model for the design of nudges. 2017.

[394] Matthew Michelson and Sofus A Macskassy. Discovering users' topics of interest on twitter: a first look. In *Proceedings of the fourth workshop on Analytics for noisy unstructured text data*, pages 73–80, 2010.

[395] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung. Digital nudging: Altering user behavior in digital environments. *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*, pages 634–648, 2017.

[396] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung. Making digital nudging applicable: The digital nudge design method. 2018.

[397] Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260, 2010.

[398] Scott SD Mitchell and Merlyna Lim. Too crowded for crowdsourced journalism: Reddit, portability, and citizen participation in the syrian crisis. *Canadian Journal of Communication*, 43(3), 2018.

[399] Tanushree Mitra and Eric Gilbert. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 49–61, 2014.

[400] Susan Mohammed and Linda C Angell. Personality heterogeneity in teams: Which differences make a difference for team performance? *Small group research*, 34(6): 651–677, 2003.

[401] Caroline Moorehead et al. Dunant's dream: War, switzerland and the history of the red cross. 1998.

[402] Richard L Moreland. Are dyads really groups? *Small Group Research*, 41(2): 251–267, 2010.

[403] Robert R Morris, Mira Dontcheva, and Elizabeth M Gerber. Priming for better performance in microtask crowdsourcing environments. *IEEE Internet Computing*, 16(5):13–19, 2012.

[404] Michael L Mortensen, Gaelen P Adam, Thomas A Trikalinos, Tim Kraska, and Byron C Wallace. An exploration of crowdsourcing citation screening for systematic reviews. *Research synthesis methods*, 8(3):366–386, 2017.

[405] Miriam Muethel, Sarah Gehrlein, and Martin Hoegl. Socio-demographic factors and shared leadership behaviors in dispersed teams: Implications for human resource management. *Human Resource Management*, 51(4):525–548, 2012.

[406] Willem J Muhren, Bartel Van de Walle, et al. Sense-making and information management in emergency response. *Bulletin of the American Society for Information Science and Technology*, 36(5):30–33, 2010.

[407] Amir Mujkanovic, David Lowe, Keith Willey, and Christian Guetl. Unsupervised learning algorithm for adaptive group formation: Collaborative learning support in remotely accessible laboratories. In *International Conference on Information Society (i-Society 2012)*, pages 50–57. IEEE, 2012.

[408] Isabel Munoz, Michael Dunn, Steve Sawyer, and Emily Michaels. Platform-mediated markets, online freelance workers and deconstructed identities. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–24, 2022.

[409] Carolin Rekar Munro and Marilyn E Laiken. Developing and sustaining high performance teams. *OD PRACTITIONER*, 35(4):63, 2003.

[410] Robert Münscher, Max Vetter, and Thomas Scheuerle. A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, 29(5): 511–524, 2016.

[411] Luis R Murillo-Zamorano, José Ángel López Sánchez, and Carmen Bueno Muñoz. Gamified crowdsourcing in higher education: A theoretical framework and a case study. *Thinking Skills and Creativity*, 36:100645, 2020.

[412] Isabel Briggs Myers. The myers-briggs type indicator: Manual (1962). 1962.

[413] Andrew Neal, Gillian Yeo, Annette Koy, and Tania Xiao. Predicting the form and direction of work role performance from the big 5 model of personality traits. *Journal of Organizational Behavior*, 33(2):175–192, 2012.

[414] George A Neuman, Stephen H Wagner, and Neil D Christiansen. The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management*, 24:28–45, 1999.

[415] Lenore Newman and Ann Dale. Homophily and agency: creating effective sustainable development networks. *Environment, development and sustainability*, 9: 79–90, 2007.

[416] Eddy S Ng and Greg J Sears. Walking the talk on diversity: Ceo beliefs, moral values, and the implementation of workplace diversity practices. *Journal of Business Ethics*, 164(3):437–450, 2020.

[417] Dong Nguyen, Rilana Gravel, Dolf Trieschnigg, and Theo Meder. " how old do

you think i am?" a study of language and age in twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 2013.

[418] Minh-Thap Nguyen and Ee-Peng Lim. On predicting religion labels in microblogging networks. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1211–1214, 2014.

[419] Jakob Nielsen. *Usability engineering*. Morgan Kaufmann, 1994.

[420] Jakob Nielsen. Usa bilitv. 2009.

[421] Peter Nijkamp and Jacques Poot. Cultural diversity: a matter of measurement. In *The economics of cultural diversity*. Edward Elgar Publishing, 2015.

[422] Melissa Nobles. Racial categorization and censuses. *Census and identity: The politics of race, ethnicity, and language in national censuses*, (1):43, 2002.

[423] Donald A Norman. *The psychology of everyday things*. Basic books, 1988.

[424] Donald A Norman and Stephen W Draper. User centered system design: New perspectives on human-computer interaction. 1986.

[425] Maria Normark. 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*, pages 81–90, 2002.

[426] Alamir Novin and Eric M Meyers. Four biases in interface design interactions. In *Design, User Experience, and Usability: Theory, Methodology, and Management: 6th International Conference, DUXU 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I 6*, pages 163–173. Springer, 2017.

[427] Chinasa Odo, Judith Masthoff, and Nigel Beacham. Group formation for collaborative learning. In *International Conference on Artificial Intelligence in Education*, pages 206–212. Springer, 2019.

[428] Chinasa Odo, Judith Masthoff, and Nigel A Beacham. Adapting online group formation to learners' conscientiousness, agreeableness and ability. In *SLLL@ AIED*, pages 1–7, 2019.

[429] Chinasa R Odo. *Adapting group formation to learners' personality traits and ability*. PhD thesis, University of Aberdeen, 2021.

[430] Ory Okolloh. Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action*, 59:65–70, 2009.

[431] Michael Boyer O'Leary and Mark Mortensen. Go (con) figure: Subgroups, imbalance, and isolates in geographically dispersed teams. *Organization science*, 21: 115–131, 2010.

[432] Anne M O'Leary-Kelly, Joseph J Martocchio, and Dwight D Frink. A review of the influence of group goals on group performance. *Academy of management journal*, 37(5):1285–1301, 1994.

[433] Judith S Olson and Wendy A Kellogg. *Ways of Knowing in HCI*, volume 2. Springer, 2014.

[434] Thomas A O'Neill and Natalie J Allen. Personality and the prediction of team performance. *European Journal of Personality*, 25:31–42, 2011.

[435] Jaderick P Pabico, Joseph Anthony C Hermocilla, John Paul C Galang, and Christine C De Sagun. Perceived social loafing in undergraduate software engineering teams. *arXiv preprint arXiv:1507.08345*, 2015.

[436] Darius Pacauskas, Risto Rajala, Mika Westerlund, and Matti Mäntymäki. Harnessing user innovation for social media marketing: Case study of a crowdsourced hamburger. *International Journal of Information Management*, 43:319–327, 2018.

[437] Aseem Pahuja and Chuan-Hoo Tan. Breaking the stereotypes: Digital nudge to attenuate racial stereotyping in the sharing economy. *Association for Information Systems AIS Electronic Library (AISeL)*, 2017.

[438] Frank-Gerald Pajonk, Burghard Andresen, Thomas Schneider-Axmann, Alexander Teichmann, Ulf Gärtner, Jürgen Lubda, Heinzpeter Moecke, and Georg von Knobelsdorff. Personality traits of emergency physicians and paramedics. *Emergency Medicine Journal*, 28(2):141–146, 2011.

[439] Stefan Palan and Christian Schitter. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[440] Leysia Palen, Starr Roxanne Hiltz, and Sophia B Liu. Online forums supporting grassroots participation in emergency preparedness and response. *Communications of the ACM*, 50(3):54–58, 2007.

[441] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5):411–419, 2010.

[442] Raja Parasuraman. Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, 43(7):931–951, 2000.

[443] Cheong Ha Park, KyoungHee Son, Joon Hyub Lee, and Seok-Hyung Bae. Crowd vs. crowd: large-scale cooperative design through open team competition. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1275–1284, 2013.

[444] Jongwon Park and JungKeun Kim. The effects of decoys on preference shifts: The role of attractiveness and providing justification. *Journal of Consumer Psychology*, 15(2):94–107, 2005.

[445] Tim F Paymans, Jasper Lindenberg, and Mark Neerincx. Usability trade-offs for adaptive user interfaces: ease of use and learnability. In *Proceedings of the 9th international conference on Intelligent user interfaces*, pages 301–303, 2004.

[446] Cecil AL Pearson. An assessment of extrinsic feedback on participation, role

perceptions, motivation, and job satisfaction in a self-managed system for monitoring group achievement. *Human relations*, 44(5):517–537, 1991.

[447] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

[448] C Carl Pegels, Yong I Song, and Baik Yang. Management heterogeneity, competitive interaction groups, and firm performance. *Strategic Management Journal*, 21 (9):911–923, 2000.

[449] John A Pelesko. *Self assembly: the science of things that put themselves together*. CRC Press, 2007.

[450] María Peña-Chilet, Gema Roldán, Javier Perez-Florido, Francisco M Ortuño, Rosario Carmona, Virginia Aquino, Daniel Lopez-Lopez, Carlos Loucera, Jose L Fernandez-Rueda, Asunción Gallego, et al. Csvs, a crowdsourcing database of the spanish population genetic variability. *Nucleic acids research*, 49(D1):D1130–D1137, 2021.

[451] Alfredo J Perez and Sherali Zeadally. A communication architecture for crowd management in emergency and disruptive scenarios. *IEEE Communications Magazine*, 57(4):54–60, 2019.

[452] Tonya M Peterson. Motivation: How to increase project team performance. *Project management journal*, 38(4):60–69, 2007.

[453] Mårten Pettersson, Dave Randall, and Bo Helgeson. Ambiguities, awareness and economy: a study of emergency service work. *Computer Supported Cooperative Work (CSCW)*, 13(2):125–154, 2004.

[454] Geoffrey Pettet, Hunter Baxter, Sayyed Mohsen Vazirizade, Hemant Purohit, Meiyi Ma, Ayan Mukhopadhyay, and Abhishek Dubey. Designing decision support systems for emergency response: Challenges and opportunities. *arXiv preprint arXiv:2202.11268*, 2022.

[455] Mark S Pfaff. Negative affect reduces team awareness: The effects of mood and stress on computer-mediated team communication. *Human Factors*, 54(4):560–571, 2012.

[456] Jonathan Pinto. *Biases and heuristics in team member selection decisions*. PhD thesis, University of Pittsburgh, 2008.

[457] David J Pittenger. Cautionary comments regarding the myers-briggs type indicator. *Consulting Psychology Journal: Practice and Research*, 57(3):210, 2005.

[458] Victoria C Plaut, Flannery G Garnett, Laura E Buffardi, and Jeffrey Sanchez-Burks. "what about me?" perceptions of exclusion and whites' reactions to multiculturalism. *Journal of personality and social psychology*, 101(2):337, 2011.

[459] Robert Plomin. Genetics and general cognitive ability. *Nature*, 402(6761):C25–C29, 1999.

[460] Marta Poblet, Esteban García-Cuesta, and Pompeu Casanovas. Crowdsourcing tools for disaster management: A review of platforms and methods. In *International Workshop on AI Approaches to the Complexity of Legal Systems*, pages 261–274. Springer, 2013.

[461] Jakob Pohlisch. An introduction to internal crowdsourcing. *Internal Crowdsourcing in Companies*, page 15, 2021.

[462] Pawel Popiel. “boundaryless” in the creative economy: assessing freelancing on upwork. *Critical Studies in Media Communication*, 34(3):220–233, 2017.

[463] Arthur E Poropat. A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, 135(2):322, 2009.

[464] Gina Potârcă and Melinda Mills. Racial preferences in online dating across european countries. *European Sociological Review*, 31(3):326–341, 2015.

[465] Liza Potts. *Social media in disaster response: How experience architects can build for participation*. Routledge, 2013.

[466] Omid Poursaeed, Tomáš Matera, and Serge Belongie. Vision-based real estate price estimation. *Machine Vision and Applications*, 29(4):667–676, 2018.

[467] J Gerard Power, Sheila T Murphy, and Gail Coover. Priming prejudice: How stereotypes and counter-stereotypes influence attribution of responsibility and credibility among ingroups and outgroups. *Human communication research*, 23: 36–58, 1996.

[468] Prolific. What is your pricing? `https://researcher-help.prolific.co/hc/en-gb/articles/360009223533-What-is-your-pricing-`, June 2021.

[469] Prolific. Prolific: Pricing. `https://www.prolific.co/pricing`, 2022.

[470] Enrico L Quarantelli. Disaster crisis management: A summary of research findings. *Journal of management studies*, 25(4):373–385, 1988.

[471] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv preprint arXiv:0909.2925*, 2009.

[472] Bahareh Rahmanian and Joseph G Davis. User interface design for crowdsourcing systems. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, pages 405–408, 2014.

[473] Arun Rai. Explainable ai: From black box to glass box. *Journal of the Academy of Marketing Science*, 48:137–141, 2020.

[474] Jorge Ramírez, Marcos Baez, Fabio Casati, and Boualem Benatallah. Understanding the impact of text highlighting in crowdsourcing tasks. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, pages 144–152, 2019.

[475] Jorge Ramírez, Burcu Sayin, Marcos Baez, Fabio Casati, Luca Cernuzzi, Boualem Benatallah, and Gianluca Demartini. On the state of reporting in crowdsourcing experiments and a checklist to aid current practices. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–34, 2021.

[476] Beatrice Rammstedt and Oliver P John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41:203–212, 2007.

[477] David G Rand. The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of theoretical biology*, 299: 172–179, 2012.

[478] Ashwini Rao, Florian Schaub, and Norman Sadeh. What do they know about me? contents and concerns of online behavioral profiles. *arXiv preprint arXiv:1506.01675*, 2015.

[479] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pages 37–44, 2010.

[480] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S Bernstein. Expert crowdsourcing with flash teams. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 75–85, 2014.

[481] Christian Reuter, Thomas Ludwig, and Volkmar Pipek. Ad hoc participation in situation assessment: Supporting mobile collaboration in emergencies. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(5):1–26, 2014.

[482] Jens Riegelsberger, Scott Counts, Shelly D Farnham, and Bruce C Philips. Sounds good to me: Effects of photo and voice profiles on gaming partner choice. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 159–162, 2006.

[483] Andreas Riener. Subliminal perception or "can we perceive and be influenced by stimuli that do not reach us on a conscious level?". In *Emotions and affect in human factors and human-computer interaction*, pages 503–538. Elsevier, 2017.

[484] Christopher Robert and Yu Ha Cheung. An examination of the relationship between conscientiousness and group performance on a creative task. *Journal of Research in Personality*, 44(2):222–231, 2010.

[485] Thomas Rocklin and William Revelle. The measurement of extroversion: A comparison of the eysenck personality inventory and the eysenck personality questionnaire. *British journal of social psychology*, 20(4):279–284, 1981.

[486] Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 321–328, 2011.

[487] Jakob Rogstadius, Maja Vukovic, Claudio A Teixeira, Vassilis Kostakos, Evangelos Karapanos, and Jim Alain Laredo. Crisistracker: Crowdsourced social media curation for disaster awareness. *IBM Journal of Research and Development*, 57(5): 4–1, 2013.

[488] Treena Eileen Rohde and Lee Anne Thompson. Predicting academic achievement with cognitive ability. *Intelligence*, 35:83–92, 2007.

[489] Jean-Pierre Rolland. The cross-cultural generalizability of the five-factor model of personality. 2002.

[490] Ruth Rosenholtz, Yuanzhen Li, Jonathan Mansfield, and Zhenlan Jin. Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 761–770, 2005.

[491] Giulio Rossetti, Luca Pappalardo, Riivo Kikas, Dino Pedreschi, Fosca Giannotti, and Marlon Dumas. Homophilic network decomposition: a community-centric analysis of online social services. *Social Network Analysis and Mining*, 6:1–18, 2016.

[492] Sebastiaan Rothmann and Elize P Coetzer. The big five personality dimensions and job performance. *SA Journal of industrial psychology*, 29:68–74, 2003.

[493] Anne C Rouse. A preliminary taxonomy of crowdsourcing. 2010.

[494] Martin Ruef, Howard E Aldrich, and Nancy M Carter. The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American sociological review*, pages 195–222, 2003.

[495] Róbert Sabo and Jakub Rajčáni. Designing the database of speech under stress. *Journal of Linguistics/Jazykovednỳ casopis*, 68(2):326–335, 2017.

[496] Abubakr Saeed, Ashiq Ali, and Hammad Riaz. Open-up or stay closed: The effect of tmt gender diversity on open innovation. *European Journal of Innovation Management*, 2023.

[497] Otto Sahlgren and Arto Laitinen. Algorithmic fairness and its limits in group-formation. In *Tethics: Conference on Technology Ethics*, pages 38–54, 2020.

[498] Eduardo Salas, James E Driskell, and Sandra Hughes. The study of stress and human performance. *Stress and human performance(A 97-27090 06-53), Mahwah, NJ, Lawrence Erlbaum Associates, Publishers, 1996,*, pages 1–45, 1996.

[499] Eduardo Salas, Katherine A Wilson, Carrie E Murphy, Heidi King, and Mary Salisbury. Communicating, coordinating, and cooperating when lives depend on it: tips for teamwork. *The Joint Commission Journal on Quality and Patient Safety*, 34(6):333–341, 2008.

[500] Eduardo Salas, Scott I Tannenbaum, Steve WJ Kozlowski, Christopher A Miller, John E Mathieu, and William B Vessey. Teams in space exploration: A new frontier for the science of team effectiveness. *Current Directions in Psychological Science*, 24(3):200–207, 2015.

[501] Niloufar Salehi and Michael S Bernstein. Hive: Collective design through network rotation. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW): 1–26, 2018.

[502] Niloufar Salehi, Andrew McCabe, Melissa Valentine, and Michael Bernstein. 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*, pages 1700–1713, 2017.

[503] Luisa Sanz-Martínez, Alejandra Martínez-Monés, Miguel L Bote-Lorenzo, Juan A Munoz-Cristóbal, and Yannis Dimitriadis. Automatic group formation in a mooc based on students' activity criteria. In *European Conference on Technology Enhanced Learning*, pages 179–193. Springer, 2017.

[504] Mike Schaekermann, Joslin Goh, Kate Larson, and Edith Law. Resolvable vs. irresolvable disagreement: A study on worker deliberation in crowd work. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–19, 2018.

[505] Eric Schenk and Claude Guittard. Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics Management*, 107:93–107, 2011.

[506] Stephen Schiffer. Ceteris paribus laws. *Mind*, 100(397):1–17, 1991.

[507] Marieke Schilpzand, David Herold, and Christina Shalley. Members' openness to experience and teams' creative performance. *Small Group Research - Small Group Res*, 41, 12 2010.

[508] Michaéla C Schippers. Social loafing tendencies and team performance: The compensating effect of agreeableness and conscientiousness. *Academy of Management Learning & Education*, 13(1):62–81, 2014.

[509] Daniel Schlagwein and Niels Bjorn-Andersen. Organizational learning with crowdsourcing: The revelatory case of lego. *Journal of the Association for Information Systems*, 15(11):3, 2014.

[510] Arjen Schmidt, Jeroen Wolbers, Julie Ferguson, and Kees Boersma. Are you ready2help? conceptualizing the management of online and onsite volunteer convergence. *Journal of Contingencies and Crisis Management*, 26(3):338–349, 2018.

[511] Christoph Schneider, Markus Weinmann, and Jan Vom Brocke. Digital nudging: guiding online user choices through interface design. *Communications of the ACM*, 61(7):67–73, 2018.

[512] Robert O Schneider. Climate change: an emergency management perspective. *Disaster Prevention and Management: An International Journal*, 2011.

[513] Silke L Schneider and AF Heath. Uncovering ethnic and cultural diversity in europe: A new classification of ethnic and cultural groups. In *Ethnicity and ancestry. 3rd International ESS Conference. Lausanne, Switzerland*, 2016.

[514] Evan Schofer, Francisco O Ramirez, and John W Meyer. The societal consequences of higher education. *Sociology of Education*, 94:1–19, 2021.

[515] Peter R Scholtes, Brian L Joiner, and Barbara J Streibel. *The team handbook.* Oriel Incorporated, 2003.

[516] Evelyn Jaffe Schreiber. Muddles and huddles: Facilitating a multicultural workforce through team management theory. *The Journal of Business Communication (1973)*, 33(4):459–473, 1996.

[517] Ella Segev. Crowdsourcing contests. *European Journal of Operational Research*, 281(2):241–255, 2020.

[518] Amartya Sen. Social choice theory. *Handbook of mathematical economics*, 3: 1073–1181, 1986.

[519] Christophe Senot, Dimitre Kostadinov, Makram Bouzid, Jérôme Picault, Armen Aghasaryan, and Cédric Bernier. Analysis of strategies for building group profiles. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 40–51. Springer, 2010.

[520] Fatma Cemile Serçe, Kathleen Swigger, Ferda Nur Alpaslan, Robert Brazile, George Dafoulas, and Victor Lopez. Online collaboration: Collaborative behavior patterns and factors affecting globally distributed team performance. *Computers in human behavior*, 27(1):490–503, 2011.

[521] Boas Shamir. Calculations, values, and identities: The sources of collectivistic work motivation. *Human relations*, 43(4):313–332, 1990.

[522] Feng Shi, Misha Teplitskiy, Eamon Duede, and James A Evans. The wisdom of polarized crowds. *Nature human behaviour*, 3(4):329–336, 2019.

[523] Shung J Shin, Tae-Yeol Kim, Jeong-Yeon Lee, and Lin Bian. Cognitive team diversity and individual team member creativity: A cross-level interaction. *Academy of Management Journal*, 55:197–212, 2012.

[524] Clay Shirky. *Here comes everybody: The power of organizing without organizations.* Penguin, 2008.

[525] Ruth Shortall, Anatol Itten, Michiel van der Meer, Pradeep K Murukannaiah, and Catholijn M Jonker. Inclusion, equality and bias in designing online mass deliberative platforms. *arXiv preprint arXiv:2107.12711*, 2021.

[526] Arturo Silva-Gálvez, RAúl Monroy, Jose E Ramirez-Marquez, and Chi Zhang. A video game-crowdsourcing approach to discover a player's strategy for problem solution to housing development. *IEEE Access*, 9:114870–114883, 2021.

[527] Chelsey S Simmons. Using catme team-maker to form student groups in a large introductory course. In *American Society for Engineering Education Southeast Section Conference*, 2015.

[528] Bernd Simon, Michael Loewy, Stefan Stürmer, Ulrike Weber, Peter Freytag, Corinna Habig, Claudia Kampmeier, and Peter Spahlinger. Collective identi-

fication and social movement participation. *Journal of personality and social psychology*, 74(3):646, 1998.

[529] Inès Skandrani-Marzouki, G Lo Monaco, and Yousri Marzouki. The effects of unconscious context on social representations: Evidence from the subliminal emotional priming paradigm. *North American Journal of Psychology*, 17(3):509–524, 2015.

[530] Alexander Smirnov, Tatiana Levashova, and Nikolay Shilov. 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)*, pages 94–101. IEEE, 2011.

[531] Kirsten A Smith, Matt Dennis, Judith Masthoff, and Nava Tintarev. A methodology for creating and validating psychological stories for conveying and measuring psychological traits. *User modeling and user-adapted interaction*, 29(3):573–618, 2019.

[532] Antonio Solanas, Rejina M Selvam, José Navarro, and David Leiva. Some common indices of group diversity: Upper boundaries. *Psychological Reports*, 111(3):777–796, 2012.

[533] Zhijun Song, Hui Zhang, and Chris Dolan. Promoting disaster resilience: Operation mechanisms and self-organizing processes of crowdsourcing. *Sustainability*, 12(5):1862, 2020.

[534] Dominic Spohr. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, 34(3):150–160, 2017.

[535] Alicia A Stachowski, Seth A Kaplan, and Mary J Waller. The benefits of flexible team interaction during crises. *Journal of Applied Psychology*, 94(6):1536, 2009.

[536] D Sandy Staples and Lina Zhao. The effects of cultural diversity in virtual teams versus face-to-face teams. *Group decision and negotiation*, 15(4):389–406, 2006.

[537] Nico Stehr. *The fragility of modern societies: Knowledge and risk in the information age*. Sage, 2001.

[538] Randy Stein and Alexander B Swan. Evaluating the validity of myers-briggs type indicator theory: A teaching tool and window into intuitive psychology. *Social and Personality Psychology Compass*, 13(2):e12434, 2019.

[539] Keri K Stephens and Brett W Robertson. 12 social media platforms and broader participation in crisis communication. *Social Media and Crisis Communication*, 2022.

[540] Flannery G Stevens, Victoria C Plaut, and Jeffrey Sanchez-Burks. Unlocking the benefits of diversity: All-inclusive multiculturalism and positive organizational change. *The journal of applied behavioral science*, 44:116–133, 2008.

[541] Brian Still and Kate Crane. *Fundamentals of user-centered design: A practical approach.* CRC press, 2017.

[542] Julia Stoyanovich, Sihem Amer-Yahia, Cameron Marlow, and Cong Yu. Leveraging tagging to model user interests in del. icio. us. In *AAAI Spring Symposium: Social Information Processing*, pages 104–109, 2008.

[543] Katherine Strater and Heather Richter. Examining privacy and disclosure in a social networking community. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 157–158, 2007.

[544] Jack W Stuster. *Bold endeavors: Lessons from polar and space exploration.* Naval Institute Press, 2011.

[545] Frederic D Stutzman, Ralph Gross, and Alessandro Acquisti. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of privacy and confidentiality*, 4(2):2, 2013.

[546] Bo-Chiuan Su, Andree Emmanuel Widjaja, and Jengchung Victor Chen. Stress in virtual team vs face-to-face team: Is working in virtual team more stressful than face-to-face team? In *STRESS IN VIRTUAL TEAM VS FACE-TO-FACE TEAM: IS WORKING IN VIRTUAL TEAM MORE STRESSFUL THAN FACE-TO-FACE TEAM?* Conference paper). Consult on, 2012.

[547] Chandrakantan Subramaniam, Hassan Ali, and Faridahwati Mohd Shamsudin. Understanding the antecedents of emergency response: a proposed framework. *Disaster Prevention and Management: An International Journal*, 2010.

[548] Jeffrey Sugerman. Using the disc® model to improve communication effectiveness. *Industrial and Commercial Training*, 2009.

[549] Sherry E Sullivan, Monica L Forret, Shawn M Carraher, and Lisa A Mainiero. Using the kaleidoscope career model to examine generational differences in work attitudes. *Career Development International*, 2009.

[550] Roshan Sumbaly, Jay Kreps, and Sam Shah. The big data ecosystem at linkedin. In *Proceedings of the 2013 acm sigmod international conference on management of data*, pages 1125–1134, 2013.

[551] Chuan Sun, Hui Li, Xiuhua Li, Junhao Wen, Qingyu Xiong, and Wei Zhou. Convergence of recommender systems and edge computing: A comprehensive survey. *IEEE Access*, 8:47118–47132, 2020.

[552] Cass R Sunstein. Nudging and choice architecture: Ethical considerations. *Yale Journal on Regulation, Forthcoming*, 2015.

[553] Cass R Sunstein. The council of psychological advisers. *Annual Review of Psychology*, 67:713–737, 2016.

[554] Cass R Sunstein. A prison of our own design: divided democracy in the age of social media. *Democratic Audit UK*, 2017.

[555] J Surowiecki. The wisdom of the crowds. new york city, 2005.

[556] Barnabas Szaszi, Anna Palinkas, Bence Palfi, Aba Szollosi, and Balazs Aczel. A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, 31 (3):355–366, 2018.

[557] Wei Tang, Ming Yin, and Chien-Ju Ho. Leveraging peer communication to enhance crowdsourcing. In *The World Wide Web Conference*, pages 1794–1805, 2019.

[558] Antonio Tapiador, Antonio Fumero, Joaquin Salvachua, and Sandra Aguirre. A web collaboration architecture. In *2006 International Conference on Collaborative Computing: Networking, Applications and Worksharing*, pages 1–4. IEEE, 2006.

[559] Monideepa Tarafdar and Jean-François Stich. Virtual work, technology and wellbeing. *The SAGE Handbook of Organizational Wellbeing, SAGE, London*, pages 159–169, 2021.

[560] Kevin Tasa, Greg J Sears, and Aaron CH Schat. Personality and teamwork behavior in context: The cross-level moderating role of collective efficacy. *Journal of Organizational Behavior*, 32(1):65–85, 2011.

[561] Leonard Tetzlaff, Florian Schmiedek, and Garvin Brod. Developing personalized education: A dynamic framework. *Educational Psychology Review*, 33(3):863–882, 2021.

[562] Richard H Thaler and Cass R Sunstein. *Nudge: Improving decisions about health, wealth, and happiness.* Penguin, 2009.

[563] Richard H Thaler, Cass R Sunstein, and John P Balz. Choice architecture. *The behavioral foundations of public policy*, 25:428–439, 2013.

[564] Edmund R Thompson. Development and validation of an international english big-five mini-markers. *Personality and individual differences*, 45(6):542–548, 2008.

[565] Peg Thoms, Keirsten S Moore, and Kimberly S Scott. The relationship between self-efficacy for participating in self-managed work groups and the big five personality dimensions. *Journal of Organizational Behavior*, 17(4):349–362, 1996.

[566] Robert Ladd Thorndike. *Applied psychometrics.* Houghton Mifflin, 1982.

[567] Nava Tintarev and Judith Masthoff. A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*, pages 801–810. IEEE, 2007.

[568] Hamid Tohidi. Teamwork productivity & effectiveness in an organization base on rewards, leadership, training, goals, wage, size, motivation, measurement and information technology. *Procedia Computer Science*, 3:1137–1146, 2011.

[569] Thomas Trainer, John R Taylor, and Christopher J Stanton. Choosing the best robot for the job: Affinity bias in human-robot interaction. In *International Conference on Social Robotics*, pages 490–501. Springer, 2020.

[570] John Tsalikis. The effects of priming on business ethical perceptions: A comparison between two cultures. *Journal of business ethics*, 131(3):567–575, 2015.

[571] Jason Tsay, Laura Dabbish, and James Herbsleb. Influence of social and technical factors for evaluating contribution in github. In *Proceedings of the 36th international conference on Software engineering*, pages 356–366, 2014.

[572] Joseph D Tucker, Suzanne Day, Weiming Tang, and Barry Bayus. Crowdsourcing in medical research: concepts and applications. *PeerJ*, 7:e6762, 2019.

[573] Bruce W Tuckman and Mary Ann C Jensen. Stages of small-group development revisited. *Group & Organization Studies*, 2(4):419–427, 1977.

[574] Ernest C Tupes and Raymond E Christal. Recurrent personality factors based on trait ratings. *Journal of personality*, 60(2):225–251, 1992.

[575] Jean M Twenge, Stacy M Campbell, Brian J Hoffman, and Charles E Lance. Generational differences in work values: Leisure and extrinsic values increasing, social and intrinsic values decreasing. *Journal of management*, 36(5):1117–1142, 2010.

[576] Unsplash. Unsplash, 2021. URL `https://unsplash.com/`.

[577] Andrew H Van de Ven, Andre L Delbecq, and Richard Koenig Jr. Determinants of coordination modes within organizations. *American sociological review*, pages 322–338, 1976.

[578] Henny van de Water, Kees Ahaus, and Robert Rozier. Team roles, team balance and performance. *Journal of Management Development*, 2008.

[579] Annelieke C Van den Berg, Sarah N Giest, Sandra M Groeneveld, and Wessel Kraaij. Inclusivity in online platforms: Recruitment strategies for improving participation of diverse sociodemographic groups. *Public Administration Review*, 80(6):989–1000, 2020.

[580] Félice van Nunspeet, Naomi Ellemers, and Belle Derks. Reducing implicit bias: How moral motivation helps people refrain from making "automatic" prejudiced associations. *Translational Issues in Psychological Science*, 1(4):382, 2015.

[581] Annelies EM Van Vianen and Carsten KW De Dreu. Personality in teams: Its relationship to social cohesion, task cohesion, and team performance. *European journal of work and organizational psychology*, 10(2):97–120, 2001.

[582] Dimiter Velev and Plamena Zlateva. Use of social media in natural disaster management. *International Proceedings of Economic Development and Research*, 39:41–45, 2012.

[583] Federica Vinella, Chinasa Odo, Ioanna Lykourentzou, and Judith Masthoff. How personality and communication patterns affect online ad-hoc teams under pressure. *Frontiers in Artificial Intelligence*, page 119, 2022.

[584] Federica Lucia Vinella, Ioanna Lykourentzou, and Judith Masthoff. Users' preferences of profiling attributes on crowdsourcing team formation systems. In

2021 16th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP), pages 1–10. IEEE, 2021.

[585] Adriana S Vivacqua and Marcos RS Borges. Taking advantage of collective knowledge in emergency response systems. *Journal of Network and Computer Applications*, 35:189–198, 2012.

[586] Svitlana Volkova, Yoram Bachrach, Michael Armstrong, and Vijay Sharma. Inferring latent user properties from texts published in social media, 2015.

[587] Karel Vredenburg, Ji-Ye Mao, Paul W Smith, and Tom Carey. A survey of user-centered design practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 471–478, 2002.

[588] Maja Vukovic. Crowdsourcing for enterprises. In *2009 congress on services-I*, pages 686–692. IEEE, 2009.

[589] Charles Waghmare. *Yammer*. Springer, 2018.

[590] Shannon L Wagner, Crystal A Martin, and Juanita A McFee. Investigating the "rescue personality". *Traumatology*, 15(3):5–12, 2009.

[591] H Jack Walker, Hubert S Feild, William F Giles, Achilles A Armenakis, and Jeremy B Bernerth. Displaying employee testimonials on recruitment web sites: effects of communication media, employee race, and job seeker race on organizational attraction and information credibility. *Journal of Applied Psychology*, 94(5):1354, 2009.

[592] H Jack Walker, Hubert S Feild, Jeremy B Bernerth, and J Bret Becton. Diversity cues on recruitment websites: Investigating the effects on job seekers' information processing. *Journal of Applied Psychology*, 97:214, 2012.

[593] Mary J Waller, Naina Gupta, and Robert C Giambatista. Effects of adaptive behaviors and shared mental models on control crew performance. *Management Science*, 50(11):1534–1544, 2004.

[594] Xinyu Wang, Zhou Zhao, and Wilfred Ng. Ustf: a unified system of team formation. *IEEE Transactions on Big Data*, 2:70–84, 2016.

[595] Szymon Wasik, Filip Fratczak, Jakub Krzyskow, and Jaroslaw Wulnikowski. Inferring mathematical equations using crowdsourcing. *Plos one*, 10(12):e0145557, 2015.

[596] LSGL Wauben, CM Dekker-van Doorn, JDH Van Wijngaarden, RHM Goossens, Robbert Huijsman, Jan Klein, and JF Lange. Discrepant perceptions of communication, teamwork and situation awareness among surgical team members. *International journal for quality in health care*, 23(2):159–166, 2011.

[597] Karl E Weick. The collapse of sensemaking in organizations: The mann gulch disaster. *Administrative science quarterly*, pages 628–652, 1993.

[598] Laurie R Weingart and Elizabeth Weldon. Processes that mediate the relationship

between a group goal and group member performance. *Human Performance*, 4: 33–54, 1991.

[599] Markus Weinmann, Christoph Schneider, and Jan Vom Brocke. Digital nudging. *Business & Information Systems Engineering*, 58(6):433–436, 2016.

[600] Yanna J Weisberg, Colin G DeYoung, and Jacob B Hirsh. Gender differences in personality across the ten aspects of the big five. *Frontiers in psychology*, page 178, 2011.

[601] David T Welsh and Lisa D Ordóñez. Conscience without cognition: The effects of subconscious priming on ethical behavior. *Academy of Management Journal*, 57 (3):723–742, 2014.

[602] Miaomiao Wen, Keith Maki, Steven Dow, James D Herbsleb, and Carolyn Rose. Supporting virtual team formation through community-wide deliberation. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–19, 2017.

[603] Eoin Whelan, Salvatore Parise, Jasper De Valk, and Rick Aalbers. Creating employee networks that deliver open innovation. *MIT Sloan Management Review*, 2011.

[604] Jacob Whitehill, Ting-fan Wu, Jacob Bergsma, Javier Movellan, and Paul Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. *Advances in neural information processing systems*, 22, 2009.

[605] Mark E Whiting, Grant Hugh, and Michael S Bernstein. Fair work: Crowd work minimum wage with one line of code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, pages 197–206, 2019.

[606] Mark E Whiting, Irena Gao, Michelle Xing, N'godjigui Junior Diarrassouba, Tonya Nguyen, and Michael S Bernstein. Parallel worlds: Repeated initializations of the same team to improve team viability. *proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–22, 2020.

[607] Jessica L Wildman, Marissa L Shuffler, Elizabeth H Lazzara, Stephen M Fiore, C Shawn Burke, Eduardo Salas, and Sena Garven. Trust development in swift starting action teams: A multilevel framework. *Group & Organization Management*, 37(2):137–170, 2012.

[608] Paula G Williams, Holly K Rau, Matthew R Cribbet, and Heather E Gunn. Openness to experience and stress regulation. *Journal of Research in Personality*, 43(5):777–784, 2009.

[609] Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah A Smith, and Frederick Liu. Crowdsourcing annotations for websites' privacy policies: Can it really work? In *Proceedings of the 25th International Conference on World Wide Web*, pages 133–143, 2016.

[610] Chi-Sum Wong and Michael A Campion. Development and test of a task level model of motivational job design. *Journal of applied Psychology*, 76(6):825, 1991.

[611] Mike Woodcock. *Team metrics: resources for measuring and improving team performance*. Routledge, 2017.

[612] Stephen Worchel and Susan L Shackelford. Groups under stress: The influence of group structure and environment on process and performance. *Personality and Social Psychology Bulletin*, 17(6):640–647, 1991.

[613] Meng-Han Wu and Alexander Quinn. Confusing the crowd: Task instruction quality on amazon mechanical turk. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 5, pages 206–215, 2017.

[614] Guo Xintong, Wang Hongzhi, Yangqiu Song, and Gao Hong. Brief survey of crowdsourcing for data mining. *Expert Systems with Applications*, 41(17):7987–7994, 2014.

[615] Zhiheng Xu, Long Ru, Liang Xiang, and Qing Yang. Discovering user interest on twitter with a modified author-topic model. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 422–429. IEEE, 2011.

[616] Lili Yang, Raj Prasanna, and Malcolm King. On-site information systems design for emergency first responders. *Journal of Information Technology Theory and Application (JITTA)*, 10:2, 2009.

[617] Shuang-Hong Yang, Bo Long, Alex Smola, Narayanan Sadagopan, Zhaohui Zheng, and Hongyuan Zha. Like like alike: joint friendship and interest propagation in social networks. In *Proceedings of the 20th international conference on World wide web*, pages 537–546, 2011.

[618] Yang Yang and Alison M Konrad. Understanding diversity management practices: Implications of institutional theory and resource-based theory. *Group & Organization Management*, 36:6–38, 2011.

[619] Virginia Yannibelli and Analía Amandi. A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. *Expert systems with applications*, 39(10):8584–8592, 2012.

[620] Weiwei Ye, Pingping Xu, Yunhua Jia, and Fen Jiang. Crowdsourcing for open innovations. *Applied Mathematics and Information Sciences*, 6(3):741–747, 2012.

[621] Jungwon Yeo, Claire Connolly Knox, and Kyujin Jung. Unveiling cultures in emergency response communication networks on social media: Following the 2016 louisiana floods. *Quality & quantity*, 52(2):519–535, 2018.

[622] Tieying Yu, Metin Sengul, and Richard H Lester. Misery loves company: The spread of negative impacts resulting from an organizational crisis. *Academy of Management Review*, 33(2):452–472, 2008.

[623] Faxi Yuan and Rui Liu. Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane matthew case study. *International journal of disaster risk reduction*, 28:758–767, 2018.

[624] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung. A survey of crowdsourcing systems. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 766–773. IEEE, 2011.

[625] Khobaib Zaamout. *A Team Composition Approach For Social Crowdsourcing Communities*. PhD thesis, University of Calgary, 2020.

[626] Stephen J Zaccaro, James Gualtieri, and David Minionis. Task cohesion as a facilitator of team decision making under temporal urgency. *Military Psychology*, 7(2):77–93, 1995.

[627] Zhi Zhai. *Technology-facilitated crowdsourcing systems*. University of Notre Dame, 2013.

[628] Haoran Zhang. A real-time optimisation for disaster-relief distribution inheterogeneous crowdsourcing. *Impact*, 2022:23–25, 2022.

[629] Shi Zhang and Gavan J Fitzsimons. Choice-process satisfaction: The influence of attribute alignability and option limitation. *Organizational behavior and human decision processes*, 77(3):192–214, 1999.

[630] Yun-Li Zhang and Chang-Qin Lu. Challenge stressor-hindrance stressor and employees' work-related attitudes, and behaviors: the moderating effects of general self-efficacy. *Acta Psychologica Sinica*, 41(06):501, 2009.

[631] Yunfeng Zhang, Rachel KE Bellamy, and Wendy A Kellogg. Designing information for remediating cognitive biases in decision-making. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 2211–2220, 2015.

[632] Wenjun Zhou, Wangcheng Yan, and Xi Zhang. Collaboration for success in crowdsourced innovation projects: Knowledge creation, team diversity, and tacit coordination. *Data Science and Analytics for Collaboration Minitrack*, 2017.

[633] Fred RH Zijlstra, Mary J Waller, and Sybil I Phillips. Setting the tone: Early interaction patterns in swift-starting teams as a predictor of effectiveness. *European Journal of Work and Organizational Psychology*, 21(5):749–777, 2012.

Summary

Crowdsourcing has become an increasingly important tool for team formation and collaboration. This thesis investigates how User-Centered Design, an iterative process that prioritizes users and their needs, can be applied to improve the efficiency and effectiveness of crowdsourcing systems for teamwork and team formation. To achieve this, we conducted a series of studies to explore the role of various factors in shaping crowd workers' behaviour and preferences in collaborative contexts.

The main findings of our research are as follows. In online team formation settings, crowd workers prefer disclosing overt traits (e.g., age, gender, topical interests) and avoid sharing sensitive information (e.g., ethnicity, depression). However, they are willing to share information regarding their personality and values, typically considered deep-level sensitive traits. Well-defined digital nudging interventions, such as a diversity progress bar, can promote diverse team formation. In contrast, subtler forms of nudging may inadvertently trigger biases working against the intended objectives.

Ad-hoc crowd teams working under pressure can benefit from systems that account for differences in personality traits, as these can influence collaboration outcomes and perceptions. Designing crowdsourcing systems for emergency response requires modelling communication tools that aid, assist, and monitor the shared load, considering the strictly cooperative roles and task- and user-dependent communication styles between collaborators. When forming teams, crowd workers tend to balance attributes between and within groups, with a preference for Openness to Experience among the Big-5 personality traits.

Based on these findings, we recommend applying a User-Centered approach to design collaborative crowdsourcing systems, considering user needs, behaviour, intents, and perceptions of digital environments. Future research should continue to explore and evaluate innovative strategies for promoting effective collaboration and team formation in crowdsourcing contexts.

Nederlandse samenvatting

Crowdsourcing is een steeds belangrijker instrument geworden voor teamvorming en samenwerking. Dit proefschrift onderzoekt hoe User-Centered Design (een iteratief proces dat prioriteit geeft aan gebruikers en hun behoeften) kan worden toegepast om de efficiëntie en effectiviteit te verbeteren van crowdsourcingsystemen voor teamwerk en teamvorming. Om dit te bereiken, hebben we een reeks onderzoeken uitgevoerd om de rol van verschillende factoren te onderzoeken die het gedrag en voorkeuren van crowdwerkers in samenwerkingscontexten beïnvloeden.

De belangrijkste bevindingen van ons onderzoek zijn als volgt: bij online teamvorming geven crowdwerkers de voorkeur aan het openbaar maken van openlijke kenmerken (bijvoorbeeld leeftijd, geslacht, interesses) en het delen van gevoelige informatie (bijvoorbeeld etniciteit, depressie) te vermijden. Ze zijn echter bereid om informatie te delen over hun persoonlijkheid en waarden, die doorgaans als gevoelig worden beschouwd.

Goed gedefinieerde digitale nudging-interventies, zoals een voortgangsbalk voor diversiteit, kunnen een diverse teamvorming bevorderen. Subtielere vormen van nudging kunnen daarentegen onbedoeld vooroordelen veroorzaken die de beoogde doelstellingen tegenwerken. Ad-hoc crowdteams die onder druk werken kunnen baat hebben bij systemen die rekening houden met verschillen in persoonlijkheidseigenschappen, omdat deze de resultaten en percepties van samenwerking kunnen beïnvloeden.

Ontwerpen van crowdsourcingsystemen voor noodhulp vereisen modellering van communicatieinstrumenten die de gedeelde last ondersteunen, assisteren en monitoren, waarbij de strikt coöperatieve rollen en taak- en gebruikersafhankelijke communicatiestijlen tussen medewerkers in acht wordt genomen. Bij het vormen van teams hebben crowdwerkers de neiging om eigenschappen tussen en binnen groepen in evenwicht te brengen, met een voorkeur voor Openheid voor Ervaring onder de Big-5-persoonlijkheidskenmerken.

Op basis van deze bevindingen raden we aan een gebruikersgerichte ontwerpbenadering toe te passen voor collaboratieve crowdsourcing-systemen, waarbij rekening wordt gehouden met de behoeften van de gebruiker, het gedrag, de intenties en hun percepties van digitale omgevingen. Toekomstig onderzoek moet innovatieve strategieën voor het bevorderen van effectieve samenwerking en teamvorming in crowdsourcingcontexten blijven onderzoeken en evalueren.

SIKS Dissertation Series

2016 01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines
02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow
03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support
04 Laurens Rietveld (VUA), Publishing and Consuming Linked Data
05 Evgeny Sherkhonov (UvA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers
06 Michel Wilson (TUD), Robust scheduling in an uncertain environment
07 Jeroen de Man (VUA), Measuring and modeling negative emotions for virtual training
08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data
09 Archana Nottamkandath (VUA), Trusting Crowdsourced Information on Cultural Artefacts
10 George Karafotias (VUA), Parameter Control for Evolutionary Algorithms
11 Anne Schuth (UvA), Search Engines that Learn from Their Users
12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems
13 Nana Baah Gyan (VUA), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach
14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization
15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments
16 Guangliang Li (UvA), Socially Intelligent Autonomous Agents that Learn from Human Reward
17 Berend Weel (VUA), Towards Embodied Evolution of Robot Organisms
18 Albert Meroño Peñuela (VUA), Refining Statistical Data on the Web
19 Julia Efremova (TU/e), Mining Social Structures from Genealogical Data
20 Daan Odijk (UvA), Context & Semantics in News & Web Search
21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground
22 Grace Lewis (VUA), Software Architecture Strategies for Cyber-Foraging Systems
23 Fei Cai (UvA), Query Auto Completion in Information Retrieval
24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior
26 Dilhan Thilakarathne (VUA), In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media
28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control
29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning
30 Ruud Mattheij (TiU), The Eyes Have It
31 Mohammad Khelghati (UT), Deep web content monitoring
32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations
33 Peter Bloem (UvA), Single Sample Statistics, exercises in learning from just one example
34 Dennis Schunselaar (TU/e), Configurable Process Trees: Elicitation, Analysis, and Enactment
35 Zhaochun Ren (UvA), Monitoring Social Media: Summarization, Classification and Recommendation
36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry
38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design
39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect
40 Christian Detweiler (TUD), Accounting for Values in Design
41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance
42 Spyros Martzoukos (UvA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora
43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice
44 Thibault Sellam (UvA), Automatic Assistants for Database Exploration
45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control
46 Jorge Gallego Perez (UT), Robots to Make you Happy
47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks
48 Tanja Buttler (TUD), Collecting Lessons Learned
49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis

50 Yan Wang (TiU), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

---

2017 01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime
02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation
03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines
04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store
05 Mahdieh Shadi (UvA), Collaboration Behavior
06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search
07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly
08 Rob Konijn (VUA), Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery
09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text
10 Robby van Delden (UT), (Steering) Interactive Play Behavior
11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment
12 Sander Leemans (TU/e), Robust Process Mining with Guarantees
13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology
14 Shoshannah Tekofsky (TiU), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior
15 Peter Berck (RUN), Memory-Based Text Correction
16 Aleksandr Chuklin (UvA), Understanding and Modeling Users of Modern Search Engines
17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution
18 Ridho Reinanda (UvA), Entity Associations for Search
19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval
20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility
21 Jeroen Linssen (UT), Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)
22 Sara Magliacane (VUA), Logics for causal inference under uncertainty
23 David Graus (UvA), Entities of Interest — Discovery in Digital Traces
24 Chang Wang (TUD), Use of Affordances for Efficient Robot Learning
25 Veruska Zamborlini (VUA), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search
26 Merel Jung (UT), Socially intelligent robots that understand and respond to human touch
27 Michiel Joosse (UT), Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors
28 John Klein (VUA), Architecture Practices for Complex Contexts
29 Adel Alhuraibi (TiU), From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"
30 Wilma Latuny (TiU), The Power of Facial Expressions
31 Ben Ruijl (UL), Advances in computational methods for QFT calculations
32 Thaer Samar (RUN), Access to and Retrievability of Content in Web Archives
33 Brigit van Loggem (OU), Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity
34 Maren Scheffel (OU), The Evaluation Framework for Learning Analytics
35 Martine de Vos (VUA), Interpreting natural science spreadsheets
36 Yuanhao Guo (UL), Shape Analysis for Phenotype Characterisation from High-throughput Imaging
37 Alejandro Montes Garcia (TU/e), WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy
38 Alex Kayal (TUD), Normative Social Applications
39 Sara Ahmadi (RUN), Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR
40 Altaf Hussain Abro (VUA), Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems
41 Adnan Manzoor (VUA), Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle
42 Elena Sokolova (RUN), Causal discovery from mixed and missing data with applications on ADHD datasets
43 Maaike de Boer (RUN), Semantic Mapping in Video Retrieval
44 Garm Lucassen (UU), Understanding User Stories - Computational Linguistics in Agile Requirements Engineering
45 Bas Testerink (UU), Decentralized Runtime Norm Enforcement
46 Jan Schneider (OU), Sensor-based Learning Support
47 Jie Yang (TUD), Crowd Knowledge Creation Acceleration
48 Angel Suarez (OU), Collaborative inquiry-based learning

2018 01 Han van der Aa (VUA), Comparing and Aligning Process Representations
02 Felix Mannhardt (TU/e), Multi-perspective Process Mining
03 Steven Bosems (UT), Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction
04 Jordan Janeiro (TUD), Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks
05 Hugo Huurdeman (UvA), Supporting the Complex Dynamics of the Information Seeking Process

06 Dan Ionita (UT), Model-Driven Information Security Risk Assessment of Socio-Technical Systems
07 Jieting Luo (UU), A formal account of opportunism in multi-agent systems
08 Rick Smetsers (RUN), Advances in Model Learning for Software Systems
09 Xu Xie (TUD), Data Assimilation in Discrete Event Simulations
10 Julienka Mollee (VUA), Moving forward: supporting physical activity behavior change through intelligent technology
11 Mahdi Sargolzaei (UvA), Enabling Framework for Service-oriented Collaborative Networks
12 Xixi Lu (TU/e), Using behavioral context in process mining
13 Seyed Amin Tabatabaei (VUA), Computing a Sustainable Future
14 Bart Joosten (TiU), Detecting Social Signals with Spatiotemporal Gabor Filters
15 Naser Davarzani (UM), Biomarker discovery in heart failure
16 Jaebok Kim (UT), Automatic recognition of engagement and emotion in a group of children
17 Jianpeng Zhang (TU/e), On Graph Sample Clustering
18 Henriette Nakad (UL), De Notaris en Private Rechtspraak
19 Minh Duc Pham (VUA), Emergent relational schemas for RDF
20 Manxia Liu (RUN), Time and Bayesian Networks
21 Aad Slootmaker (OU), EMERGO: a generic platform for authoring and playing scenario-based serious games
22 Eric Fernandes de Mello Araújo (VUA), Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks
23 Kim Schouten (EUR), Semantics-driven Aspect-Based Sentiment Analysis
24 Jered Vroon (UT), Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots
25 Riste Gligorov (VUA), Serious Games in Audio-Visual Collections
26 Roelof Anne Jelle de Vries (UT),Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology
27 Maikel Leemans (TU/e), Hierarchical Process Mining for Scalable Software Analysis
28 Christian Willemse (UT), Social Touch Technologies: How they feel and how they make you feel
29 Yu Gu (TiU), Emotion Recognition from Mandarin Speech
30 Wouter Beek (VUA), The "K" in "semantic web" stands for "knowledge": scaling semantics to the web

---

2019 01 Rob van Eijk (UL),Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification
02 Emmanuelle Beauxis Aussalet (CWI, UU), Statistics and Visualizations for Assessing Class Size Uncertainty
03 Eduardo Gonzalez Lopez de Murillas (TU/e), Process Mining on Databases: Extracting Event Data from Real Life Data Sources
04 Ridho Rahmadi (RUN), Finding stable causal structures from clinical data
05 Sebastiaan van Zelst (TU/e), Process Mining with Streaming Data
06 Chris Dijkshoorn (VUA), Nichesourcing for Improving Access to Linked Cultural Heritage Datasets
07 Soude Fazeli (TUD), Recommender Systems in Social Learning Platforms
08 Frits de Nijs (TUD), Resource-constrained Multi-agent Markov Decision Processes
09 Fahimeh Alizadeh Moghaddam (UvA), Self-adaptation for energy efficiency in software systems
10 Qing Chuan Ye (EUR), Multi-objective Optimization Methods for Allocation and Prediction
11 Yue Zhao (TUD), Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs
12 Jacqueline Heinerman (VUA), Better Together
13 Guanliang Chen (TUD), MOOC Analytics: Learner Modeling and Content Generation
14 Daniel Davis (TUD), Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses
15 Erwin Walraven (TUD), Planning under Uncertainty in Constrained and Partially Observable Environments
16 Guangming Li (TU/e), Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models
17 Ali Hurriyetoglu (RUN),Extracting actionable information from microtexts
18 Gerard Wagenaar (UU), Artefacts in Agile Team Communication
19 Vincent Koeman (TUD), Tools for Developing Cognitive Agents
20 Chide Groenouwe (UU), Fostering technically augmented human collective intelligence
21 Cong Liu (TU/e), Software Data Analytics: Architectural Model Discovery and Design Pattern Detection
22 Martin van den Berg (VUA),Improving IT Decisions with Enterprise Architecture
23 Qin Liu (TUD), Intelligent Control Systems: Learning, Interpreting, Verification
24 Anca Dumitrache (VUA), Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing
25 Emiel van Miltenburg (VUA), Pragmatic factors in (automatic) image description
26 Prince Singh (UT), An Integration Platform for Synchromodal Transport
27 Alessandra Antonaci (OU), The Gamification Design Process applied to (Massive) Open Online Courses
28 Esther Kuindersma (UL), Cleared for take-off: Game-based learning to prepare airline pilots for critical situations
29 Daniel Formolo (VUA), Using virtual agents for simulation and training of social skills in safety-critical circumstances
30 Vahid Yazdanpanah (UT), Multiagent Industrial Symbiosis Systems
31 Milan Jelisavcic (VUA), Alive and Kicking: Baby Steps in Robotics
32 Chiara Sironi (UM), Monte-Carlo Tree Search for Artificial General Intelligence in Games
33 Anil Yaman (TU/e), Evolution of Biologically Inspired Learning in Artificial Neural Networks
34 Negar Ahmadi (TU/e), EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES
35 Lisa Facey-Shaw (OU), Gamification with digital badges in learning programming

36 Kevin Ackermans (OU), Designing Video-Enhanced Rubrics to Master Complex Skills
37 Jian Fang (TUD), Database Acceleration on FPGAs
38 Akos Kadar (OU), Learning visually grounded and multilingual representations

2020 01 Armon Toubman (UL), Calculated Moves: Generating Air Combat Behaviour
02 Marcos de Paula Bueno (UL), Unraveling Temporal Processes using Probabilistic Graphical Models
03 Mostafa Deghani (UvA), Learning with Imperfect Supervision for Language Understanding
04 Maarten van Gompel (RUN), Context as Linguistic Bridges
05 Yulong Pei (TU/e), On local and global structure mining
06 Preethu Rose Anish (UT), Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support
07 Wim van der Vegt (OU), Towards a software architecture for reusable game components
08 Ali Mirsoleimani (UL),Structured Parallel Programming for Monte Carlo Tree Search
09 Myriam Traub (UU), Measuring Tool Bias and Improving Data Quality for Digital Humanities Research
10 Alifah Syamsiyah (TU/e), In-database Preprocessing for Process Mining
11 Sepideh Mesbah (TUD), Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models
12 Ward van Breda (VUA), Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment
13 Marco Virgolin (CWI), Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming
14 Mark Raasveldt (CWI/UL), Integrating Analytics with Relational Databases
15 Konstantinos Georgiadis (OU), Smart CAT: Machine Learning for Configurable Assessments in Serious Games
16 Ilona Wilmont (RUN), Cognitive Aspects of Conceptual Modelling
17 Daniele Di Mitri (OU), The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences
18 Georgios Methenitis (TUD), Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems
19 Guido van Capelleveen (UT), Industrial Symbiosis Recommender Systems
20 Albert Hankel (VUA), Embedding Green ICT Maturity in Organisations
21 Karine da Silva Miras de Araujo (VUA), Where is the robot?: Life as it could be
22 Maryam Masoud Khamis (RUN), Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar
23 Rianne Conijn (UT), The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging
24 Lenin da Nóbrega Medeiros (VUA/RUN), How are you feeling, human? Towards emotionally supportive chatbots
25 Xin Du (TU/e), The Uncertainty in Exceptional Model Mining
26 Krzysztof Leszek Sadowski (UU), GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization
27 Ekaterina Muravyeva (TUD), Personal data and informed consent in an educational context
28 Bibeg Limbu (TUD), Multimodal interaction for deliberate practice: Training complex skills with augmented reality
29 Ioan Gabriel Bucur (RUN), Being Bayesian about Causal Inference
30 Bob Zadok Blok (UL), Creatief, Creatiever, Creatiefst
31 Gongjin Lan (VUA), Learning better – From Baby to Better
32 Jason Rhuggenaath (TU/e), Revenue management in online markets: pricing and online advertising
33 Rick Gilsing (TU/e), Supporting service-dominant business model evaluation in the context of business model innovation
34 Anna Bon (UM), Intervention or Collaboration? Redesigning Information and Communication Technologies for Development
35 Siamak Farshidi (UU), Multi-Criteria Decision-Making in Software Production

2021 01 Francisco Xavier Dos Santos Fonseca (TUD),Location-based Games for Social Interaction in Public Space
02 Rijk Mercuur (TUD), Simulating Human Routines: Integrating Social Practice Theory in Agent-Based Models
03 Seyyed Hadi Hashemi (UvA), Modeling Users Interacting with Smart Devices
04 Ioana Jivet (OU), The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning
05 Davide Dell'Anna (UU), Data-Driven Supervision of Autonomous Systems
06 Daniel Davison (UT), "Hey robot, what do you think?" How children learn with a social robot
07 Armel Lefebvre (UU), Research data management for open science
08 Nardie Fanchamps (OU), The Influence of Sense-Reason-Act Programming on Computational Thinking
09 Cristina Zaga (UT), The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play
10 Quinten Meertens (UvA), Misclassification Bias in Statistical Learning
11 Anne van Rossum (UL), Nonparametric Bayesian Methods in Robotic Vision
12 Lei Pi (UL), External Knowledge Absorption in Chinese SMEs
13 Bob R. Schadenberg (UT), Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning
14 Negin Samaeemofrad (UL), Business Incubators: The Impact of Their Support
15 Onat Ege Adali (TU/e), Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm
16 Esam A. H. Ghaleb (UM), Bimodal emotion recognition from audio-visual cues
17 Dario Dotti (UM), Human Behavior Understanding from motion and bodily cues using deep neural networks

18 Remi Wieten (UU), Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks
19 Roberto Verdecchia (VUA), Architectural Technical Debt: Identification and Management
20 Masoud Mansoury (TU/e), Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems
21 Pedro Thiago Timbó Holanda (CWI), Progressive Indexes
22 Sihang Qiu (TUD), Conversational Crowdsourcing
23 Hugo Manuel Proença (UL), Robust rules for prediction and description
24 Kaijie Zhu (TU/e), On Efficient Temporal Subgraph Query Processing
25 Eoin Martino Grua (VUA), The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications
26 Benno Kruit (CWI/VUA), Reading the Grid: Extending Knowledge Bases from Human-readable Tables
27 Jelte van Waterschoot (UT), Personalized and Personal Conversations: Designing Agents Who Want to Connect With You
28 Christoph Selig (UL), Understanding the Heterogeneity of Corporate Entrepreneurship Programs

---

2022 01 Judith van Stegeren (UT), Flavor text generation for role-playing video games
02 Paulo da Costa (TU/e), Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey
03 Ali el Hassouni (VUA), A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare
04 Ünal Aksu (UU), A Cross-Organizational Process Mining Framework
05 Shiwei Liu (TU/e), Sparse Neural Network Training with In-Time Over-Parameterization
06 Reza Refaei Afshar (TU/e), Machine Learning for Ad Publishers in Real Time Bidding
07 Sambit Praharaj (OU), Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics
08 Maikel L. van Eck (TU/e), Process Mining for Smart Product Design
09 Oana Andreea Inel (VUA), Understanding Events: A Diversity-driven Human-Machine Approach
10 Felipe Moraes Gomes (TUD), Examining the Effectiveness of Collaborative Search Engines
11 Mirjam de Haas (UT), Staying engaged in child-robot interaction, a quantitative approach to studying preschoolers' engagement with robots and tasks during second-language tutoring
12 Guanyi Chen (UU), Computational Generation of Chinese Noun Phrases
13 Xander Wilcke (VUA), Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented Knowledge
14 Michiel Overeem (UU), Evolution of Low-Code Platforms
15 Jelmer Jan Koorn (UU), Work in Process: Unearthing Meaning using Process Mining
16 Pieter Gijsbers (TU/e), Systems for AutoML Research
17 Laura van der Lubbe (VUA), Empowering vulnerable people with serious games and gamification
18 Paris Mavromoustakos Blom (TiU), Player Affect Modelling and Video Game Personalisation
19 Bilge Yigit Ozkan (UU), Cybersecurity Maturity Assessment and Standardisation
20 Fakhra Jabeen (VUA), Dark Side of the Digital Media - Computational Analysis of Negative Human Behaviors on Social Media
21 Seethu Mariyam Christopher (UM), Intelligent Toys for Physical and Cognitive Assessments
22 Alexandra Sierra Rativa (TiU), Virtual Character Design and its potential to foster Empathy, Immersion, and Collaboration Skills in Video Games and Virtual Reality Simulations
23 Ilir Kola (TUD), Enabling Social Situation Awareness in Support Agents
24 Samaneh Heidari (UU), Agents with Social Norms and Values - A framework for agent based social simulations with social norms and personal values
25 Anna L.D. Latour (UL), Optimal decision-making under constraints and uncertainty
26 Anne Dirkson (UL), Knowledge Discovery from Patient Forums: Gaining novel medical insights from patient experiences
27 Christos Athanasiadis (UM), Emotion-aware cross-modal domain adaptation in video sequences
28 Onuralp Ulusoy (UU), Privacy in Collaborative Systems
29 Jan Kolkmeier (UT), From Head Transform to Mind Transplant: Social Interactions in Mixed Reality
30 Dean De Leo (CWI), Analysis of Dynamic Graphs on Sparse Arrays
31 Konstantinos Traganos (TU/e), Tackling Complexity in Smart Manufacturing with Advanced Manufacturing Process Management
32 Cezara Pastrav (UU), Social simulation for socio-ecological systems
33 Brinn Hekkelman (CWI/TUD), Fair Mechanisms for Smart Grid Congestion Management
34 Nimat Ullah (VUA), Mind Your Behaviour: Computational Modelling of Emotion & Desire Regulation for Behaviour Change
35 Mike E.U. Ligthart (VUA), Shaping the Child-Robot Relationship: Interaction Design Patterns for a Sustainable Interaction

---

2023 01 Bojan Simoski (VUA), Untangling the Puzzle of Digital Health Interventions
02 Mariana Rachel Dias da Silva (TiU), Grounded or in flight? What our bodies can tell us about the whereabouts of our thoughts
03 Shabnam Najafian (TUD), User Modeling for Privacy-preserving Explanations in Group Recommendations
04 Gineke Wiggers (UL), The Relevance of Impact: bibliometric-enhanced legal information retrieval
05 Anton Bouter (CWI), Optimal Mixing Evolutionary Algorithms for Large-Scale Real-Valued Optimization, Including Real-World Medical Applications
06 António Pereira Barata (UL), Reliable and Fair Machine Learning for Risk Assessment
07 Tianjin Huang (TU/e), The Roles of Adversarial Examples on Trustworthiness of Deep Learning

08 Lu Yin (TU/e), Knowledge Elicitation using Psychometric Learning
09 Xu Wang (VUA), Scientific Dataset Recommendation with Semantic Techniques
10 Dennis J.N.J. Soemers (UM), Learning State-Action Features for General Game Playing
11 Fawad Taj (VUA), Towards Motivating Machines: Computational Modeling of the Mechanism of Actions for Effective Digital Health Behavior Change Applications
12 Tessel Bogaard (VUA), Using Metadata to Understand Search Behavior in Digital Libraries
13 Injy Sarhan (UU), Open Information Extraction for Knowledge Representation
14 Selma Čaušević (TUD), Energy resilience through self-organization
15 Alvaro Henrique Chaim Correia (TU/e), Insights on Learning Tractable Probabilistic Graphical Models
16 Peter Blomsma (TiU), Building Embodied Conversational Agents: Observations on human nonverbal behaviour as a resource for the development of artificial characters
17 Meike Nauta (UT), Explainable AI and Interpretable Computer Vision – From Oversight to Insight
18 Gustavo Penha (TUD), Designing and Diagnosing Models for Conversational Search and Recommendation
19 George Aalbers (TiU), Digital Traces of the Mind: Using Smartphones to Capture Signals of Well-Being in Individuals
20 Arkadiy Dushatskiy (TUD), Expensive Optimization with Model-Based Evolutionary Algorithms applied to Medical Image Segmentation using Deep Learning
21 Gerrit Jan de Bruin (UL), Network Analysis Methods for Smart Inspection in the Transport Domain
22 Alireza Shojaifar (UU), Volitional Cybersecurity
23 Theo Theunissen (UU), Documentation in Continuous Software Development
24 Agathe Balayn (TUD), Practices Towards Hazardous Failure Diagnosis in Machine Learning
25 Jurian Baas (UU), Entity Resolution on Historical Knowledge Graphs
26 Loek Tonnaer (TU/e), Linearly Symmetry-Based Disentangled Representations and their Out-of-Distribution Behaviour
27 Ghada Sokar (TU/e), Learning Continually Under Changing Data Distributions
28 Floris den Hengst (VUA), Learning to Behave: Reinforcement Learning in Human Contexts
29 Tim Draws (TUD), Understanding Viewpoint Biases in Web Search Results

---

2024 01 Daphne Miedema (TU/e), On Learning SQL: Disentangling concepts in data systems education
02 Emile van Krieken (VUA), Optimisation in Neurosymbolic Learning Systems
03 Feri Wijayanto (RUN), Automated Model Selection for Rasch and Mediation Analysis
04 Mike Huisman (UL), Understanding Deep Meta-Learning
05 Yiyong Gou (UM), Aerial Robotic Operations: Multi-environment Cooperative Inspection & Construction Crack Autonomous Repair
06 Azqa Nadeem (TUD), Understanding Adversary Behavior via XAI: Leveraging Sequence Clustering to Extract Threat Intelligence
07 Parisa Shayan (TiU), Modeling User Behavior in Learning Management Systems
08 Xin Zhou (UvA), From Empowering to Motivating: Enhancing Policy Enforcement through Process Design and Incentive Implementation
09 Giso Dal (UT), Probabilistic Inference Using Partitioned Bayesian Networks
10 Cristina-Iulia Bucur (VUA), Linkflows: Towards Genuine Semantic Publishing in Science
11 withdrawn
12 Peide Zhu (TUD), Towards Robust Automatic Question Generation For Learning
13 Enrico Liscio (TUD), Context-Specific Value Inference via Hybrid Intelligence
14 Larissa Capobianco Shimomura (TU/e), On Graph Generating Dependencies and their Applications in Data Profiling
15 Ting Liu (VUA), A Gut Feeling: Biomedical Knowledge Graphs for Interrelating the Gut Microbiome and Mental Health
16 Arthur Barbosa Câmara (TUD), Designing Search-as-Learning Systems
17 Razieh Alidoosti (VUA), Ethics-aware Software Architecture Design
18 Laurens Stoop (UU), Data Driven Understanding of Energy-Meteorological Variability and its Impact on Energy System Operations
19 Azadeh Mozafari Mehr (TU/e), Multi-perspective Conformance Checking: Identifying and Understanding Patterns of Anomalous Behavior
20 Ritsart Anne Plantenga (UL), Omgang met Regels
21 Federica Vinella (UU), Crowdsourcing User-Centered Teams
22 Zeynep Ozturk Yurt (TU/e), Beyond Routine: Extending BPM for Knowledge-Intensive Processes with Controllable Dynamic Contexts

Publications

Publications included in the thesis

[**Chapter 3 is extended from:**]Vinella, F. L., Lykourentzou, I., & Masthoff, J. (2021, November). Users' Preferences of Profiling Attributes on Crowdsourcing Team Formation Systems. In 2021 16th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP) (pp. 1-10). IEEE.

[**Chapter 4 is extended from:**] Vinella, F., Mosch, R., Lykourentzou, I., & Masthoff, J. (2022, July). The Impact of Digital Nudging Techniques on the Formation of Self-Assembled Crowd Project Teams. In Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (pp. 265-275).

[**Chapter 5 is extended from:**] Vinella, F. L., Odo, C., Lykourentzou, I., & Masthoff, J. (2022). How Personality and Communication Patterns Affect Online ad-hoc Teams Under Pressure. Frontiers in Artificial Intelligence, 5, 818491.

[**Chapter 6 is extended from:**] Vinella, F. L., Koppelaar, S., & Masthoff, J. (2022, July). Forming Teams of Learners Online in a User as Wizard Study with Openness, Conscientiousness, and cognitive Ability. In Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (pp. 283-292).

Other publications

Vinella, F. L., Hu, J., Lykourentzou, I., & Masthoff, J. (2022). Crowdsourcing team formation with worker-centered modeling. Frontiers in Artificial Intelligence, 5, 818562.

Vinella, F. L., Lykourentzou, I., & Papangelis, K. (2020, July). Motivational Principles and Personalisation Needs for Geo-Crowdsourced Intangible Cultural Heritage Mobile Applications. In Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization (pp. 362-369).

Lykourentzou, I., Vinella, F. L., Ahmed, F., Papastathis, C., Papangelis, K., Khan, V. J., & Masthoff, J. (2021). Self-organizing teams in online work settings. arXiv preprint arXiv:2102.07421.

Hannewijk, B., Vinella, F. L., Khan, V. J., Lykourentzou, I., Papangelis, K., & Masthoff, J. (2020). Capturing the city's heritage on-the-go: Design requirements for mobile crowdsourced cultural heritage. Sustainability, 12(6), 2429.

Saccardi, I., Islakoglu, D. S., Neerincx, A., & Vinella, F. L. (2022). Symbiotic Child Emotional Support with Social Robots and Temporal Knowledge Graphs. arXiv preprint arXiv:2205.13229.

Curriculum Vitæ

Federica Lucia Vinella

February 16th, 1990 Putignano, Italy.

Education

2017–2018	Masters in Computer Science Edinburgh Napier University, Edinburgh, UK *Thesis:* DRUG NAME DETECTION WITH DEEP LEARNING *Supervisors:* Dimitra Gkatzia *Graduation Grade: 8.2*
2009–2011	Bachelors in Fine Arts Accademia di Brera, Milano, Italia *Thesis:* CONTROCULTURA UNDERGROUND: I Movimenti Sotterranei Del Ventesimo Secolo *Supervisor:* Marcella Anglani *Graduated with Cum Laude*
2003–2008	Liceo delle Scienze Sociali Leonardo da Vinci, Noci (BA), Italia *Graduation Grade: 9.1*

Work

2019–present	PhD Researcher at the Human Centered Computing Group Utrecht University
2012–2017	Self-Employed - freelancer and Furniture Designer Federica Designs, Vinella & Krupa, Edinburgh, UK

Acknowledgements

In 2019, I traded the windswept majesty of Edinburgh for the charming canals of Utrecht, embarking on a bold quest to become a researcher. Leaving behind seven and a half years in Scotland – a tartan threaded with memories of mountains, castles, lochs, friendships, and folk songs – I embarked on a new Dutch chapter. Five years of PhD research, a global pandemic, the joys of teaching, and the grit of scientific exploration have flown by. This transfigurative journey has sculpted me into someone profoundly different. I am writing this acknowledgement section because I owe the life achievement of completing a PhD to myself as much as those who stood by my side, believed it was possible, and told me never to give up. I had the unyielding support of my family, many colleagues, old and new friends, and the love of many pets, including Spip, Kliko, Peppino (you are very missed!) and the guinea pigs at Castellum. Being near pets kept me going during the most challenging moments. This thesis would not have been possible without their pretty eyes, soft fur, and purrs.

I sincerely thank my supervisors, Judith Masthoff and Ioanna Lykourentzou, as well as the dissertation committee members Julita Vassileva, Marko Tkalcic, Remco Veltkam, Lynda Hardman, and Anja Volk; they provided much wisdom that helped me chart my course. A heartfelt thank you to my paranymphs, Isabella Saccardi and Onuralp Ulusoy. Your support throughout the lead-up to my Defense was invaluable. Your positive energy was a soothing balm that calmed my nerves and helped me persevere. I would also like to express my deepest gratitude to all my colleagues at Utrecht University. A big thank you goes to the constant figure in my PhD career, Anouk Neerincx. You were the first to welcome me with open arms just a few months into my PhD journey, and you readily embarked on this adventure with me. Thank you for your staunch support, even when things got tough. There was an instant connection between us, and I knew from the beginning that our friendship would be a constant source of strength. I want to acknowledge my colleagues who have enriched academia with their skills, talent, and genuine kindness. This acknowledgement goes to those colleagues who understood me and shared their experiences with open hearts. I thank my colleagues from the Human-Centered Computing group. I want to thank Susanne Poeller for bringing a genuine sense of humanity and honesty into the office environment. Although you joined the research group only recently, we already share many meaningful memories, and I look forward to more. I want to thank Francisca Pessanha de Meneses Ribeiro dos Reis for being open to discussion and being present every step of the way. Although you may not see it this way, you are a role model for many; thank you for being sensitive, caring, and human. Lennart Herlaar, thank you for being my teaching role model. All the sound teaching practices I learned are thanks to you. Your professionalism was a beacon that made me see teaching in a very different light. Also, thank you for the Pluim Award; it would have been impossible to get it without you.

I use this acknowledgement section to celebrate the positive impact many at Utrecht University have left throughout my PhD journey. Thank you to Almila Akdag, Maartje de Graaf, Hanna Hauptmann, Eelco Herder, Christof van Nimwegen, Imke de Jong, Lukas Arts, Karlijn Dinnissen, Kazi Haque, Anouk van Kasteren, Mathijs Langezaal, Bjorn van Zwol, Shah Noor Khan, Robin Cromjongh, Anneloes Meijer, Federica Giusti, Marit Bentvelzen, Zerrin Yumak, Egon van den Broek, Joeroen Ooge, Marloes Vredenborg, Julie Pivin-Bachler, Simone Ooms, Robbert Jan Beun, Albert Salah, Gizem Sogancioglu, Metehan Doyran, Heysem Kaya, Bilgecag Aydogdu, Evanthia

Dimara, Frans Wiering, Remco Veltkamp, Dionysis Alexandridis, Christina Verver-van Ek, Johan Jeuring, Gabrielle Keller, Sander Bakkers, Arno Siebes, Xixi Lu, Vinicius Stein Dani, Jan Martijn van der Werf, Kees van Deemter, Albert Gatt, Dong Nguyen, Marijn Schraagen, Eduardo Calo', Shihan Wang, Enas Khwaileh, Duygu Islakoglu, Masha Ghanavatinasab, Vahid Shahrivari, Ro Jefferson, Till Miltzow, Sukanya Pandey, Gerdard Tel, Stoop Laurens, Sibel Telli, Ernestina Hauptfeld, Sylvia Zweers. Every one of you has contributed in some way to making this long academic journey memorable, shaping me not only as a scholar but also as a person. If I forgot to mention anyone, please know I am grateful for your help.

Jan Pieter, I sincerely thank you. Your inexhaustible patience and steady calm were my lifeline during moments of doubt. I couldn't have persevered without your essential support – your conscientiousness, tenacity, and keen mind were invaluable. You were a partner, a sounding board, and a second reviewer, ensuring my writing met the highest standards. To my incredible family, especially Mum, Dad, and Gianni, thank you for your constant emotional refuge. Your daily WhatsApp calls and frequent visits nourished my spirit. I extend my heartfelt appreciation to all my other family members, cousins, uncles, and aunties, whose unwavering belief in me fueled my journey. Your positivity and constant encouragement meant the world. Celebrating this milestone with all of you in Puglia, my childhood home, fills me with joy. Imagine us – feasting on scrumptious Barese focaccia, indulging in refreshing pistachio gelato, and basking in the sun with the sparkling blue sea as our backdrop. It will be the perfect way to mark this new chapter.

To my incredible friends scattered across Italy, Scotland, France, and beyond, I feel fortunate to have your support wherever you may be. Sumaita Tahseen, thank you for being my fearless travel companion, a constant source of inspiration, and a true embodiment of feminine strength. Our bike adventures have been a breath of fresh air, offering the perfect escape from routine and bringing freedom and peace into my life. We still have so much of the world to explore together! Alisa Rieger, your open ears and willingness to share experiences have been invaluable. Our fun outings in Den Haag were bright spots in my week, adding a welcome spice to the routine. Aga Koundourakis and Helene Dres, a decade of friendship and counting! I cherish our bond and know it will continue to grow stronger. Rosa Mosch, working with such a brilliant mind like yours has been a true pleasure. Your intelligence and enthusiasm made our collaboration a delightful and enriching experience.

Ingrid Roijenga, your constant encouragement to stay active and positive has been a lifeline throughout my PhD journey. Thank you for reminding me of the importance of well-being. A big thank you goes to my childhood friend Lorella Bagorda. Although we haven't been able to spend time recently, I know you are by my side every step of the way. Thank you to my fashionable friends Ramona Viglianisi and Sumru for bringing sparkle and fun into my life. A big thank you to all those people I met in Utrecht who have decided to stay in my life. Friendship is indeed the rainbow in someone's cloud. Alex Lewis, Francine and Helen, Ingrid Zukerman, Dennis van der Wal, Olaf Takens, Hiskia and Daniel Hopman, Klaas and Andrea Zuidersma, Veronica from Taranto, and many more. Looking back, this PhD journey has been an unforgettable experience. Each supportive person, each encouraging conversation, and each shared laugh has contributed to the vibrant picture before me. This thesis is not just a culmination of my work but a testament to the collective power of human connection. From the bottom of my heart, thank you all.

I would also like to acknowledge the unexpected sources of strength I found throughout my PhD journey. Literature, pop culture, and video games gave me much-needed escapes, inspiration, and moments of pure joy. From the fantastical worlds of Harry Potter and Middle-earth to the strategic depths of Civilization and the fascinating interstellar explorations of Star Trek, these fictional realms offered solace, fueled creativity and reminded me of the importance of play. A heartfelt thank you to these unexpected companions on my academic adventure.