Artificial intelligence methods to support people management in organisations

by

EWA DOMINIKA ANDREJCZUK

A dissertation presented in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in Computer Science

Advisors:
Prof. Carles Sierra
Prof. Juan Antonio Rodriguez-Aguilar

Tutor:
Prof. Carles Sierra

Universitat Autònoma de Barcelona
Departamento de Sistemas Multiagente

2018

Abstract

Organisations have shifted from work arranged around individual jobs to team-based work structures. A new generation of solutions for organisations must give support to team management by encouraging team effectiveness and introducing automation. In this dissertation, we tackle several different problems that are connected to team management in organisations. In particular, we contribute by proposing a people management workflow that addresses the problems connected to team composition as well as problems of accurate employee evaluation and task performance evaluation.

First, we review the literature on team composition and formation from both the organisational psychology and computer science perspectives and we explore the connection between individuals' attributes and team performance as well as the cross fertilization opportunities between those fields.

Second, we review the most prominent tools to measure individuals' attributes, as these measures are necessary inputs for team composition processes. In particular, we describe the dominant approaches in Organisational Psychology, Industrial Psychology and Human Resources and summarise they main findings to measure individual personality and competences.

Third, we use our findings to propose a model to predict team performance given a task and based on individuals' attributes (i.e. competences, personality and gender). We define the Synergistic Team Composition Problem (STCP) as the problem of finding a team partition constrained by size so that each team, and the whole partition of employees into teams, is balanced in terms of individuals' competences, personality and gender. We propose two different algorithms to solve this problem: an optimal algorithm called STCPSolver that is effective for small instances of the problem, and an approximate algorithm called SynTeam that provides high-quality, but not necessarily optimal solutions. We present empirical results that we obtained when analysing student performance. Our results show the benefits of a more informed team composition that exploits individuals' competences, personalities and gender.

Fourth, we devise an algorithm called Collaborative Judgment (CJ) to fairly evaluate individuals' and teams' outcomes once tasks are performed. In particular, we want to diminish the importance of biases in the evaluation process by allowing evaluators to assess their peers, namely other evalutors. Our empirical results show the benefits of more informed assessment aggregation method.

Acknowledgements

This incredible journey comes to an end. I can honestly say that doing a PhD is one of the most wonderful things I have done in my life. It also happens to be one of the toughest ones. Along the way I encountered dead-ends, hidden forks, snakes and it was dark at times. But I kept pushing and I stayed focused to earn something eternal: the title of doctor. I did it indeed but... I could not have done it without a countless support of people around me.

I am forever indebted to my academic supervisors, Juan A. Rodriguez-Aguilar and Carles Sierra. In my journey towards this degree, you were my teachers, my friends, my inspiration, my role models and my pillars of support. Thank you for guiding me through the process of becoming a scientist. Thank you for your inspiration, encouragement, criticism, and, above all, patience and for always having time for me. You kept me focused and ensured I stay on course and do not deviate from the core of my research. Without your able guidance, this thesis would not have been possible and I shall eternally be grateful.

To Alvaro Ortin, Matthew Walsh, Luis Artiles, Harald Messemer and an entire crew of Enzyme and CMT SL for receiving me with open arms. Wihout you this would not have been possible and your support and many discussions inspired the entire course of this thesis.

I extend much gratitude to all teachers and students who helped with validating my algorithms, especially to Carme Roig, Sonia Roman, Rosa Duran Fonta and Angel Vidal Escales.

A great part of this dissertation has been developed at a very special place, the Artificial Intelligence Research Institute (IIIA-CSIC). Thanks to all my fellow colleagues in this adventure. With some of them the paths crossed for a short period of time, but they became an important part of my day-to-day life in the institute. To Paula, Kemo and Tito for suffering with me. To Albert, Borja, Ramon, Sergi and Toni for endless discussions in the lab. To Javier and Xavi for sharing the PhD wisdom. To Oguz for singing "In the Jungle" when I most needed it. I would also like to thank Andrew, Angels, Bruno, Christian, Filippo, Jesus, Lissette, Maite, Nardine, Patricia, Pere, Roberto, Felix, Ramon, and the other staff members, who were there when I needed a hand.

I also want to thank all those who have supported me throughout this long journey. To my awesome polish girls, Martyna, Ruda and Katia. Thanks for all the visits, laughs, countless evenings with wine and keeping me (in)sane. To

the people that made Barcelona feel like home. To Francys for being mi morena favorita. To Alicia for Spanish lessons free-of-charge and frenetic Ibiza nights. To Isaac for sharing my mambo passion. To Alex for messing with sandwiches and keeping me safe. To Manel for hikes, especially the ones to come. To so many people I met here and I failed to mention.

Finally, above all, I want to thank my family. To my mother Anna and my father Ryszard who encouraged me to dream big and to my brother, Grzegorz, my sister-in-law, Agnieszka, my aunt Ela and my grandma Lola. Thank you for always being there for me, and for your constant support. To my nephew, Filip, hopefully you want to read this thesis one day (although you don't have to).
I love you all.

– Ewa Andrejczuk

This work has been partially supported by an Industrial PhD scholarship from the Generalitat de Catalunya (DI-060).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Societies continually evolve and change demands creating the need for new products and services. Companies are often forced to make changes to stay competitive. A constant process of change has become part of every organisation as changes provoke other changes. It is stability, not change, that becomes the exception, and organisations are forced to find their way in those circumstances. One of the adaptations to effectively manage change is to process decision-making down in the organisation hierarchy [Ouye, 2011]. Companies can no longer adapt by just reducing costs. Decision making has to be fast and cannot wait to go up and down the management chain. Therefore, it becomes crucial for company's success to have competent professionals at all levels. Hence, in this thesis we look at people management from two different levels, i.e. individual employees as well as teams and organisations.

When looking for an employee, most organisations have a standard recruitment process: human resources start by reviewing résumés, move on to phone screening, then schedule face-to-face interviews with the most promising candidates, to finally draw on various tests to measure applicants aptitudes, personality and competences [Bateson et al., 2013]. Recent research shows that about 76% of companies with more than one hundred employees perform competence, behavioral and personality tests for recruitment [Chamorro-Premuzic, 2015]. Yet, to the best of our knowledge, once a professional is hired, organisations do not use the information about new employees collected during the recruitment process. The competences of employees, if measured at all, are collected through self-assessment tests or derived from periodic appraisals performed within an organisation [Barth and de Beer, 2017]. However, self-assessments and current appraisal processes suffer from a significant amount of bias. Additionally, it is not clear how the collected data is used besides the purpose of comparing year-to-year individual improvement for a pay raise and layoff processes.

Looking at the individual alone is not enough. Individuals are to some extent bounded by the norms of the groups they belong to [Ouye, 2011]. Within companies and conglomerates, as well as in government agencies and schools, teams are now the fundamental unit of organisation [Duhigg, 2016]. Teams provide a structure and means of bringing together people with a suitable mix of individual attributes. This can encourage the exchange of ideas, creativity, motivation and job satisfaction and can actually extend individuals' capabilities. In turn, a suitable team can improve the overall productivity in the organisation, and the quality of the performed tasks. However, sometimes a team may work less effectively than initially expected due to several reasons: a bad balance of team members' capacities, incorrect team dynamics, lack of communication, or difficult social situations of team members. While respectful disagreements can be productive, some personality differences can lead to disruptive conflicts. The opposite behaviour is as much harmful, making team members quietly accept initial ideas without questioning and a discussion of alternatives. Teams also might face difficulties when some team members do not contribute as much as others. Yet, to the best of our knowledge, teams in organisations are mostly handcrafted without giving much thought about team attributes and synergies. As far as we can tell, there are no computational methods to compose teams for given tasks that are widely used in organisations. Additionally, once a team performs a task, the information on task success or failure is not included when re-evaluating competences of employees.

In this thesis we tackle several different problems that are connected to people management in organisations. First, there have been many methods developed to measure individuals' attributes (mainly competences and personality). However, having so many different methods makes it difficult to select the most appropriate ones. Therefore, we aim at reviewing most prominent tools to measure individuals' attributes, their construct validity issues, their popularity and their pros and cons. Second, even though the majority of organisations nowadays organise their work around teams [Kozlowski and Ilgen, 2006], to the best of our knowledge, there is no single method accepted widely by organisations to compose teams. Also in research, team composition and formation problems are of interest to many fields of science, primarily to organisational psychology, but also of computer science. However, both fields have evolved separately disregarding the results of the other field. Therefore, we plan to review the literature on team composition and formation from both fields to explore the connection between individuals' attributes and team performance as well as the cross fertilization opportunities between those fields. Once we know the state-of-the-art, we wish to offer a method for predicting team performance given a task and based on the gathered information about individuals. Currently, to the best of our knowledge, the individual information collected in organisations is safely stored but rarely re-used. Next, we aim at designing methods to compose effective teams given a list of employees within a single department. In particular, we want to focus on both finding "the best" team for a task. Moreover, we aim at composing a set of teams for a given task so that each team is balanced in individuals' attributes with the purpose of increasing the performance of whole department. Third, organisations are in a constant need of evaluating the performance of both individuals and teams, however currently used methods in organisations allow for a significant amount of bias. Without an accurate performance measures, we neither can evaluate competences fairly nor we are able to correctly predict team performance. Hence, we intent to devise an algorithm to evaluate fairly individuals' and teams' outcomes once tasks are performed. In particular, we want to diminish the importance of biases in the evaluation process by allowing employees to comment on the fairness of appraisals.

The remaining of this chapter is organised as follows. In Section 1.1 we discuss further the motivation that led to this work. In Section 1.2 we present the research questions for this thesis. Next, in Section 1.3 we discuss the vision on team management process inspiring this thesis. In Section 1.4 we highlight the contributions of this dissertation and we elaborate on the structure of this dissertation.

1.1 Motivation

In this section we delve into the motivation that led us to this research. We focus on two subcategories that are crucial for teamwork in organisations. First, we start by identifying current pitfalls and shortcoming in the management of *Individuals* within organisations. Second, we explain the challenges that organisations face when managing teams as the fundamental organisational element.

1.1.1 Individuals

Individuals are at the heart of everything that is accomplished in organisations. It is nearly impossible for the organisation to advance in meaningful ways unless their employees are competent and motivated. Accurate appraisal of individuals' attributes (competences, motivation, stress rate, performance etc.) is the best way to gain insight into how useful for a company its employees really are. Also, the insights from competence appraisals can be used as an input for team processes. They can help in composing effective teams, predict a team success for a given task as well as discover the necessity for adding a new employee in a team. *Yet, when reviewing the literature on competence and performance appraisals, there appears to be no one single best process that is widely used in organisations.*

When hiring a new employee, individual attributes are difficult to measure as appraisal sources are limited, though there are various tools that can be used to assess individuals' attributes. The most common method is an interview, where a candidate is asked a set of questions to assess her level of competences and personality needed for the position, her fit to a company's culture, etc [Gusdorf, 2008]. Typically, interview questions require the candidate to give real examples of past projects, and it is based on the theory that past behavior is a good predictor of future behavior [Deb, 2006]. The interview is actually a verbal test, however the results are subject to interpretation by the interviewer. Hence, the outcome of the interview has a considerable potential for error, depending on the interviewer's own personal bias. Moreover, a candidate might gloss over her past, leaving space for assessment mistake. Human resources typically ask a new employee to provide biographical data, the results of academic and development programs, training and seminar certificates as well as recommendation letters. All these sources can be used as part of an assessment to measure the competence level of the new employee. Additionally, some organisations ask individuals to fill in cognitive ability tests or perform work sample tests that are similar to tasks that she can encounter during the real job. These tests are based on the premise that the best predictor of future behavior is observed behavior under similar situations [Deb, 2006]. Collecting all this information is helpful in getting a whole picture of the candidate, however it still does not guarantee the candidate's success in the real job tasks.

Once the employee works in the organisation for some time, appraisals are the means to an individual's career development by helping identify and set goals for the employee, recognize progress over time, identify problem areas and motivate. Typically, an appraisal is a real number called *performance rating* that is obtained from a converted and/or aggregated collection of assessments [Shaout and Yousif, 2014]. In the conventional performance appraisal or review process, a manager periodically (normally annually) writes her assessment on the performance of a reporting employee (ibid.). This is the simplest and least costly solution, although allowing for significant biases. These biases can go both ways — employees can benefit or be punished by the manager personal likes and dislikes [Buckingham, 2011]. Firstly, managers tend to remember the most recent events instead of analyzing the entire year's performance. Secondly, the importance of an initial impression might heavily influence an appraisal, irrespectively of subsequent performance. Finally, a personal bias can come from a manager's views about race, nationality, gender, religion, age, disability, hair colour, intelligence, etc.

When it comes to organisations, knowing the talent they have makes it easier to discover recruitment necessities, to build competent teams for given tasks, to estimate the probability of projects' success, etc. Therefore, some companies try to reduce the biases by collecting assessments from various sources. One of the most widely used tools is the 180/360 multi-rater feedback method [Barth and de Beer, 2017], where feedback of peers, self-assessments or even direct reports or clients are included to help evaluate an employee's true competences. *Assessments are subjective by nature, although having multiple sources makes the aggregation less subjective.* Typically, these assessments are aggregated with respect to a type of reviewer (direct managers, peers, direct reports, the employee herself, etc.), using a simple or weighted mean of all given assessments (like in systems such as Hudson (uk.hudson.com), Success Factors (successfactors.com), Halogen Software (halogensoftware.com), Appraisal-smart (appraisal-smart.com), WLH Consulting (wlhconsulting.com) and many more). This solution is still not ideal for a number of reasons. Firstly, multi-rater appraisal focuses on rating a person's performance in a given period of time. These appraisals are too broad and too subjective, making the collected data biased [Buckingham, 2011]. Secondly, the number of reviewers required for this assessment method needs to be relatively high (for instance the experts from Halogen software recommend to use between eight and fifteen assessors for development focused evaluations [Saba, 2017]). Therefore, some organisations find it too expensive to collect and process that amount of data every year or half a year.

The cost versus quality trade-off makes it extremely difficult to choose one single method for employee appraisal. Nevertheless, the 180/360 process could be less costly if it was smartly introduced in an organisation. For instance, code reviews or integration tests can be a good opportunity to evaluate programming, architecture or design skills of engineers. *Also, the information of potential biases could be included by allowing reviewers to comment on the assessments of others. However, to the best of our knowledge, current technologies used in organisations do not make usage of opinions expressed about assessments.* We think that this kind of information is very important as it can be key to build the reputation of assessors. A bad assessor can be detected by the assessing community if they were allowed to simply express their opinions about the bad assessor. Actually, in many social networks this kind of information is collected ("was this recommendation useful to you?"), and presented to users. However, how the sites use this information to rank recommendations is never clearly explained if it is used at all.

Having accurate unbiased assessment is essential not only for business management but also for computer science, particularly in the area of multiagent systems. There, agents' individual performance is key for team and coalition composition [Osman et al., 2013]. Team formation and coalition formation are crucial for many applications related to multiagent cooperation, e.g. RoboCup rescue team [Nair et al., 2003; Ramchurn et al., 2010], Unmanned Aerial Vehicles (UAVs) operations [Haque et al., 2013], or team formation in social networks [Lappas et al., 2009] to name just a few. Both team formation and coalition formation focus on assembling the *best* possible group of agents (be it either a team or a coalition) to accomplish some tasks of interest given some limited resources. Hence, it is crucial for these algorithms to count on an assessment of the *expected capabilities* of the agents to recruit.

Given this background, in this thesis we review currently used competence and performance assessment methods and we propose an evaluation algorithm based on the collective opinion of assessors.

1.1.2 Teams and Organisations

The latter part of the 20$^{\text{th}}$ and the beginning of the 21$^{\text{st}}$ centuries have witnessed a significant transformation from work organised around individual jobs to team-based work structures together with a focus on organisational efficiency [Kozlowski and Bell, 2013]. This is due to the increasing complexity of tasks, which in many cases cannot be performed by single individuals [Ramezan, 2011]. The complex tasks need the concourse of several people composing teams. *Yet, even though much research in different fields focused on the predictors of team performance, most organisations handcraft their teams ignoring the insights coming from the literature.*

Team composition has attracted researchers from different fields, mainly from organisational psychology and industrial psychology, but also from computer science, especially in the area of multiagent systems [Chiocchio et al., 2015; Osman et al., 2013]. Nevertheless, research on team composition and team formation in computer science and organisational psychology also has evolved separately. On the one hand, multiagent literature has typically disregarded significant organisational psychology findings, with the exception of several recent, preliminary attempts (such as [Alberola et al., 2016; Farhangian et al., 2015a; Hanna and Richards, 2015]) focusing on algorithms that help automate team formation and composition. On the other hand, the organisational psychology literature has mainly focused on empirically investigating the factors that influence team performance to develop heuristics that help organisations handcraft their teams. *Despite the common research interests shared by the multiagent and organisational psychology literature, to the best of our knowledge there has been no effort in the literature to bridge the knowledge produced by both research disciplines.*

In organisational and industrial psychology, we distinguish between two approaches to team composition, that is The Individual Attributes Approach and The Team Balance Approach.

The Individual Attributes Approach is based on the presumption that, when it comes to predicting a team's performance, some individual attributes matter more than others. Hence, considerable work in those fields has focused on identifying what attributes are important and how to use these attributes to build effective teams [Arnold and Randall, 2010; Mount et al., 1998; Schmidt and Hunter, 1998; White, 1984]. *These factors include competences, experiences, age and gender as well as personality.* Numerous studies [Arnold and Randall, 2010; Mount et al., 1998; White, 1984] underline the importance of personality traits or types in team composition and formation. The most popular personality tests used to explore this approach are: the Myers-Briggs Type Indicator (MBTI) [Myers et al., 1998] and the Five Factor Model (aka FFM [Costa and McCrae, 1992] or "Big Five"[Goldberg, 1990]). The MBTI consists of four dichotomous dimensions that are represented on a binary scale, that is: Extraversion / Introversion (EI), Sensing / Intuition (SN), Thinking / Feeling (TF), Judging / Perceiving (JP). These dimensions are designed to indicate how individuals perceive the world and make decisions [Myers et al., 1998]. The Five Factor Model uses five broad dimensions to describe different aspects of human personality, that is: Extraversion, Agreeableness, Consciousness, Emotional Stability and Openness to Experience [Costa and McCrae, 1992]. In the Individual Attributes approach research examines attributes on a one-at-a-time basis. It also typically suggests that some individuals are simply better working in teams than others. *We believe taking the Individual Approach is counter-intuitive as some people may work well together, while others may not and it rather depends on the compatibility between team members.*

Henceforth, some researchers in organisational psychology focus on the Team Balance Approach where they try to understand which team member attributes are best in terms of the configuration that they compose. Here, the question is not whether the team's mean on a given, single variable affects team performance (as in the research outlined above). This general approach explores if team members complement each other based on the particular composition of several attributes associated with each team member. *Surprisingly, research examining the Team Balance Approach has been very limited.*

The team composition theories that take this perspective vary depending on individual attributes they focus on (e.g. experience, personality, level of skill, or gender, among others) [West, 2012b]. Schutz [Schutz, 1958] proposed the Fundamental Interpersonal Relations Orientations (FIRO) theory based on the idea that there are three human needs relevant to teamwork that need to be balanced within a team. These needs are: need for inclusion, need for control and need for affection. According to Schutz, the most effective teams are those that are composed of individuals whose scores on different needs vary substantially. Another theory proposed by Belbin emphases the importance of *roles* in the team composition [Aritzeta et al., 2007]. In essence, Belbin claims that there are nine team roles required that need to be balanced for a team to be effective: Implementer, Resource Investigator, Co-ordinator, Teamworker, Plant, Shaper, Monitor Evaluation, Completer Finisher and Specialist. Each person can have multiple roles. In order for a team to be most effective, all roles should be covered. More recent research findings [Wilde, 2009, 2013] suggest that both a diverse personality profile of team members and a balanced gender distribution, positively influence the effectiveness of a team. Here, effectiveness is understood as the probability of goal achievement while performing problem-solving tasks. Therefore, they propose a novel team composition method that is based on the Post-Jungian Personality Theory [Wilde, 2013]. The questionnaire measures the same dimensions as MBTI, although it uses *numerical* data collected by the questionnaire, instead of binary measure as used in MBTI.

Team composition and formation are critical issues also for co-operative multiagent systems. There, the question behind team composition and formation is how to create a multiagent system as a group of heterogeneous agents (such as humans, robots or software agents) and how to organise their activities. MAS research has widely acknowledged competences as important for performing tasks of different nature [Anagnostopoulos et al., 2012; Chen et al., 2015; Crawford et al., 2016; Okimoto et al., 2015; Peleteiro et al., 2015; Rangapuram et al., 2013]. However, the majority of the approaches represent capabilities of agents in a Boolean way (i.e., an agent either has a required skill or not). This is a simplistic way to model an agent's set of capabilities as it ignores any skill degree. In real life, capabilities are not binary since every individual shows different performances for each competence. MAS literature has typically disregarded the vast experience of Organisational Psychology about assessing individuals, as well as heuristic knowledge about team composition (besides recent, preliminary attempts, like [Farhangian et al., 2015a; Hanna and Richards, 2015]).

1.2 Open Research Questions

Related to the topic of individual attributes assessments, and team assessments and composition, there are many questions that can be addressed. In this section, we introduce the specific challenges to be tackled in this thesis.

- ***Question 1:*** *Are there unexplored cross-fertilization ideas between the Computer Science and the Organisational Psychology fields when it comes to team composition and formation?*

There is a need to provide an integrative perspective on team composition, team formation and their relationship with team performance. Thus, we review the contributions in both the computer science literature and the organisational psychology literature dealing with these topics. We argue that computer science and organisational psychology have followed rather disparate approaches when it comes to team composition and team formation. However, some similarities and differences can be drawn. Henceforth, we want to review current findings from organisational psychology and computer science, to analyse and compare the strengths and weaknesses of their contributions, and to identify research gaps and opportunities by bringing together the knowledge of the two research strands on team composition and formation. Our analysis also pursues to identify cross-fertilisation opportunities that help both disciplines benefit from one another. Given the volume of existing literature, this review is not intended to be exhaustive. Instead, we plan to focus on the most significant contributions in both fields together with recent contributions that break new ground to spur innovative research.

- ***Question 2:*** *Can we devise a method to predict a single team performance better than experts?*

Predicting team performance is an important issue in team–based organisations, especially in environments that require teams to be constantly created and dismantled, such as software development, scientific experiments, or the classroom. If future team performance could be predicted, it would be beneficial for human resource planning, training and recruitment. Additionally, if we could forecast future team performance, we could use this information to compose effective teams. Henceforth, in this thesis, we want to explore how individual attributes influence team performance. In particular, we aim at checking if given individuals' attributes, such as competences, personality and gender, we are able to build a model to predict team performance better than experts.

- ***Question 3:*** *Is there a method to split an organisation into teams so that they work overall better than the teams composed by experts?*

Teams are one way for organisations to gather insights from members, and to provide employees with a sense of involvement in the pursuit of organisational goals. Teams exist within an organisation and interact with one another within this organisation. Thus, having one best team is not enough. Organisations focus on team improvement, so that all of them work effectively. Henceforth, in this thesis, we are interested in composing *a set of effective teams.* We aim at checking if the overall performance of teams composed by our method is better than performance of teams composed by the experts.

- ***Question 4:*** *Can we diminish the importance of biases when assessing individual and team performance?*

There are various ways to assess individual and team performance. As mentioned in Section 1.1, the assessment typically consists of the opinion of one assessor, or a simple or weighted average of the opinions of several assessors. As a part of this thesis, we are interested in designing an assessment method that is able to identify incompetent reviewers and reduce their contribution to the final performance rating.

1.3 Team Management: a vision

A new generation of solutions for business management in organisations must give support to team management by introducing automation to accelerate decision-making. Given the motivation described in section 1.1, this research addresses several different problems crucial for organisations as we identify in section 1.1. First, we overview existing methods to evaluate individuals' competences and personality and discuss issues connected to these methods, so that the most appropriate tool can be chosen. Second, we offer a model to predict team's performance based on the attributes of team members and with respect to a task that is assigned to the team. This model can be used to compose a single effective team. Third, we propose two different algorithms to divide employees from given department into teams based on individuals' attributes (i.e. competences, personality and gender) so that the overall performance of the department is high.

Before we describe further our contributions, we want to discuss the team management organisational workflow that we intend to automate. The workflow shown in figure 1.1 is intended to provide a general framework for team management within organisations.

We identify the main roles and processes present in this team management workflow as follows:

1. **Roles:**
    - **Employees:** We have a pool of employees to form teams within an organisation.
    - **Human Resources:** Human resources are responsible for the initial collection and assessment of employees' attributes, that is, personal data, competences and personality.

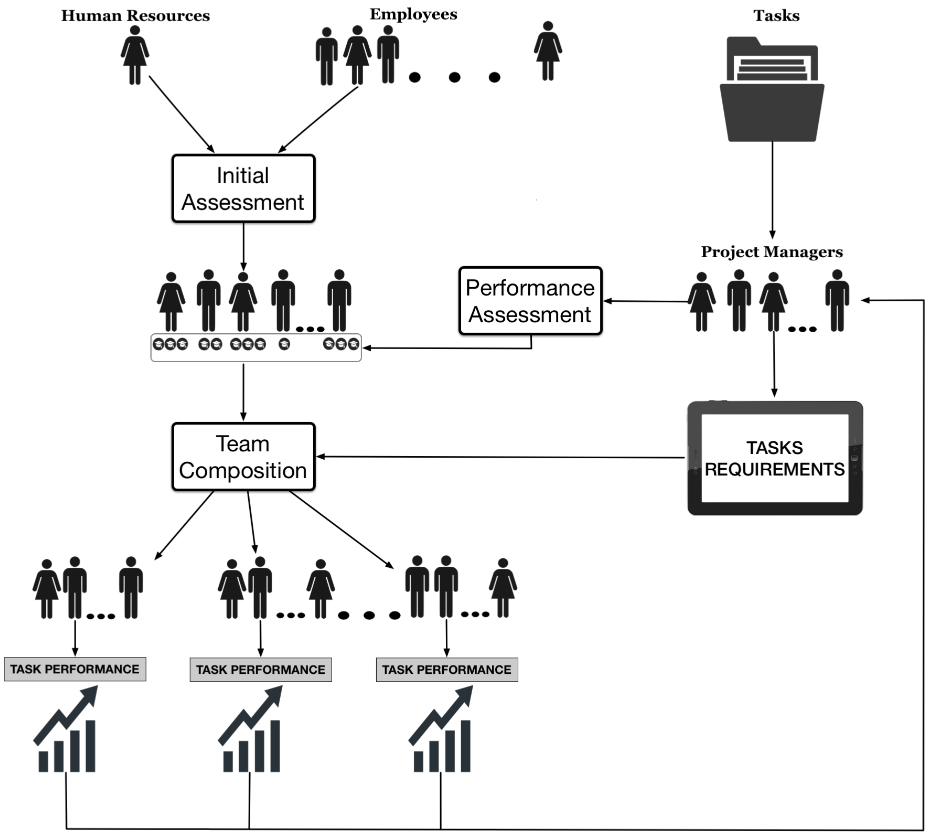


Figure 1.1: The team management organisational workflow.

- **Project Managers:** The role of project managers is to specify task requirements (such as the number of employees needed and competences required) and evaluate team and individual performance upon task completion. Additionally, project managers assist in the performance of the tasks observing if the requirements of the task are defined well and do not change with time. If the requirements change, project managers may be obliged to change the team composition adding or subtracting team members.

2. **Processes:**

- **Initial Assessment:** This process is manual or semi-automatic and consists of the collection and assessment of input data for an organisational flow, that is, personal data of employees (competences, personality, gender etc). When it comes to the initial assessment of competences and personality, evaluation sources can vary. It can come from cognitive assessments, self-assessment questionnaires, work samples, employee background, competence-based and behavioural interviews, assessment centres, peer-assessments and many more.

- **Team Composition:** Given a list of task requirements specified by project managers a team composition process either composes one *best* team or divides a department of employees into teams so that each team is both competent and team members work well together with the purpose of increasing overall performance of that department.
- **Performance Assessment:** Once tasks are performed, data goes back to project managers who assess the outcomes of the tasks. Based on this data, project managers write their opinions about individual contributions in the task performance assessing also individuals' competences. Employees can also assess their peers' competences.

In detail, the complete flow goes as follows. The process starts with the Human Resources Department, a set of employees that are subject to team composition and a set of tasks that need to be performed. First, Human Resources perform *The Initial Assessment* of personal attributes (such as competences, personality and gender) of all employees. At the same time, Project Managers specify Requirements for tasks. Once the initial assessment is done and task requirements specified, *The Team Composition* process is triggered composing teams for each task. Finally, upon the completion of tasks, the outcomes go back to project managers who use this data to do the *Performance Assessment.* The feedback provided by project managers is further ahead employed for *The Team Composition* process.

1.4 Contributions & Guide To The Thesis

In this dissertation we contribute with algorithms providing support to the team management problems.

In Chapter 2 we discuss the background and research work relevant to this thesis. This thesis is related to peer assessments, as well as team composition and formation from both computer science and organisational psychology perspective. The literature related to those subjects is large. Therefore, we introduce some of the most recent and related work in these areas. This chapter also addresses ***Question 1*** by providing an integrative perspective on team composition, team formation and their relationship with team performance. We review the knowledge produced by both the computer science literature and the organisational psychology literature dealing with these topics. Our purpose is twofold. First, we aim at identifying the strengths and weaknesses of the contributions made by these two diverse bodies of research. Second, we pursue to identify cross-fertilisation opportunities that help both disciplines benefit from one another. To the best of our knowledge there has been no attempt to integrate and compare the contributions provided by those two fields.

In Chapter 3 we analyze methods used to assess individual attributes that are shown to be correlated to team performance. We discuss existing methods that can be used to measure personalities and competences of individuals within an organisation. We also consider pros and cons of each method. The conclusions of this chapter are presented in Section 6.5.

Question 2 is addressed in Chapter 4, where we propose a model to predict performance of a single team given a task. The model serves also as a purely automatic method to compose teams based on individuals' attributes. In detail, key factors influencing team performance are competences and personality of team members. Hence, we present a computational model to evaluate *proficiency* and *congeniality* of teams based on individuals' personalities and their competences to perform tasks of different nature. With this purpose, we extend Wilde's post-Jungian method for team composition, which solely employs individuals' personalities and gender. To the best of our knowledge, this is a first computational model to compose teams based on individuals' competences, their personality and their gender. In order to answer ***Question 2***, we perform the experiments in an educational scenario. In current school practice, teachers group students according to their own, manual method based on the knowledge about students, their competences, background and social situation. Therefore, we pitch our automated team composition model with the team composition performed by teachers. In detail, we compare both team composition models in terms of how well they predict team performance. Our empirical results show a gain up to 50% in prediction accuracy with respect to teachers. Finally, we discuss the implications of this work as well as the potential usage in the team composition problems in Section 4.6

In Chapter 5 we address ***Question 3:***. First, we define the Synergistic Team Formation Problem (STCP) as the problem of finding a team partition constrained by size whose *synergistic value* is maximal. We regard our team composition problem as a particular type of set partition problem. Namely, we are interested in a split of a set of individuals into teams so that each team, and the whole partition of agents into teams, is balanced in terms of competences, personality, gender and team size. For this purpose, we use the model presented in Chapter 4. To the best of our knowledge, there is no attempt in the literature to solve our problem. Henceforth, in addition to presenting the synergistic team formation problem, we contribute by developing algorithms for solving the problem. These algorithms are potentially useful for any organisation that faces the need to optimise their problem solving teams (e.g. a classroom, a company, a research unit). The first algorithm is based on an ILP formulation and its solution by a commercial ILP solver. While for small instances this approach is rather successful, this is not case for larger problem instances. Hence, we also develop a heuristic for the STCP, called SynTeam, that is meaningful for organisations and classrooms. Our computational results show that the heuristic approach underpins a powerful algorithm for the synergistic team composition problem. For instance, for 45 agents and team size equal to 5, we observe that SynTeam is able to provide very good solutions (quality ratio of over 95%) in less than 3 seconds, while ILP solver needs approximately 700 seconds to come up with a first, low-quality solution. In order to reach optimality ILP solver requires 233 times the time required by SynTeam. We also present empirical results that we obtained when analysing student performance in order to answer ***Question 3***. We benchmark our team composition method with the current school practice. We perform two different experiments in education scenario with the total of 252 students to show the effectiveness of our approach. In the first study, the relative improvement of teams composed by SynTeam vs teams composed by traditional method is equal to 29.2%. In the second study, the relative improvement is equal to 25.3%. Our results show the benefits of a more informed team composition that exploits individuals' competences, personalities and gender. We discuss this work in subsection 5.5.

Chapter 6 addresses ***Question 4***. We introduce a new ranking algorithm, called *Collaborative Judgement (CJ)* to evaluate: (1) individuals competences which deals with issues raised in Chapter 3, (2) Outcomes of teams' task performance. Collaborative Judgments algorithms takes into account *peer opinions* of agents and/or humans on objects (e.g. products, exams, papers) as well as *peer judgements* over those opinions. The combination of these two types of information has not been studied in the literature in order to produce object rankings. The algorithm is of general purpose, however in order to test it, we decided to apply *Collaborative Judgement* to the use case of scientific paper assessment and we validate it over simulated data. We compare CJ with the standard algorithm used in Conference Management Systems (like Confmaster or Easychair) that weighs opinions with the assessors' self-assessments. We call this simple algorithm *Self-Assessment Weighted Algorithm* (SAWA). The results show that CJ algorithm outperforms SAWA, as it is much more resilient to biased reviewers. As a matter of fact, as opposed to SAWA that treats all reviewers equally, CJ is designed to detect biased reviewers and diminish the importance of their opinions by the usage of the reputation measure. We observe that CJ's gains become larger than 20% and statistically significant for percentages of good reviewers between 20% and 80%. These results answer ***Question 4***. The conclusions of this work are presented in 6.5.

The material contained in this thesis has been published and/or presented as the following articles (to be corrected before deposition of the thesis):

- **Don't Leave Anyone Behind: Imposing Team Performance through Diversity**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra, Carme Roig, Yolanda Parejo-Romero; Under revision for the The 48th Annual Frontiers in Education Conference (FIE), 2018.

- **Heterogeneous Teams for Homogeneous Performance**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra, Filippo Bistaffa, Christian Blum; Under revision for the 27th International Joint Conference on Artificial Intelligence and the 23rd European Conference on Artificial Intelligence (IJCAI-ECAI), 2018.

- **Solving The Synergistic Team Formation Problem (Extended abstract)**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra, Filippo Bistaffa, Christian Blum; Proceedings of the 17th Conference on Autonomous Agents and MultiAgent Systems. International Foundation for Autonomous Agents and Multiagent Systems (AAMAS), 2018.

- **Collaborative Rankings**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra; *Fundamenta Informaticae 158*, 2018, p.277–295;

- **Synergistic Team Composition (Extended abstract)**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra, Carme Roig; Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems. International Foundation for Autonomous Agents and Multiagent Systems (AAMAS), 2017.

- **Synergistic Team Composition.** (2017); Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carme Roig, Carles Sierra; *First International Workshop on Teams in Multiagent Systems (TEAMAS)*, May, 2017;

- **Congenial Teamsourcing.**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carme Roig, Carles Sierra; *First International Workshop on Teams in Multiagent Systems (TEAMAS)*, May, 2017;

- **The Composition and Formation of Effective Teams. Computer Science meets Organisational Psychology** (IN PRESS); Ewa Andrejczuk, Rita Berger, Juan A. Rodríguez-Aguilar, Carles Sierra, Víctor Marín-Puchades; Manuscript accepted by The Knowledge Engineering Review Journal;

- **A Concise Review on Multiagent Teams: Contributions and Research Opportunities**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra; *Proceedings of the 14th European Conference on Multiagent Systems (EUMAS)*, 2016;

- **Optimising Congenial Teams**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carme Roig, Carles Sierra; *International Workshop on Optimization in multiagent systems (OPTMAS)*, May, 2016;

- **Collaborative assessments in on-line classrooms**; Nardine Osman, Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra; The proceedings of 7th International Workshop on Collaborative Agents Research & Development: CARE for Digital Education; Springer, Cham, 2016. p. 97-116;

- **Collaborative Judgement**; Ewa Andrejczuk, Juan A. Rodríguez-Aguilar, Carles Sierra; *Proceedings of 18th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2015), pp. 631-639*; Springer International Publishing, 2016;

Chapter 2

Background and Related Work

In this chapter, we discuss the background and the literature related to this thesis. We start from introducing the fundamental terminology used in this thesis in section 2.1. Next, we move on to presenting an analysis of the state of the art for the peer assessment in the field of multiagent systems in section 2.2. Later on, in section 2.3 we move on to give an integrative perspective on team composition, team formation and their relationship with team performance. In order to do this, we review the contributions in the computer science literature and the organisational psychology literature dealing with these topics. Given the volume of existing literature, our review is not intended to be exhaustive. Instead, we have preferred to focus on the most significant contributions together with recent contributions that break new ground to spur innovative research.

2.1 Background

In this section we introduce the fundamental terminology used in this thesis.

2.1.1 Team Vocabulary

In this thesis we refer to:

1. Team Composition as the process of deciding which agents will be part of a team,

2. Team Formation as the process of learning by agents to work together in a team and through this learning decide the roles and internal organisation of a team,

3. Teamwork as the process of performing a task by a composed and formed team.

While there is a common understanding of teamwork within both OP and CS, the scientists do not agree on the notion of team formation. In computer science it is mostly understood as the process of deciding which agents will be a part of a team (here called team composition). Our definition of team formation is in line with the organisational psychology literature [Kozlowski and Bell, 2013, p.16].

Another discrepancy between the computer science and the organisational psychology literature is the notion of skill and competence. Typically in computer science all kinds of agents' competences are called skills, while in OP the definition is more complex. In OP a prominent conceptualization of competence was given by Roe [Roe, 2002b, p.195]. He defines competence as "a learned ability to adequately perform a task, duty or role". Following his definition competences "integrate knowledge, skills, personal values, and attitudes and are build on knowledge and skills and are acquired through work experience and learning by doing" [Bartram and Roe, 2005]. Hence, competences include abilities and behaviours, as well as knowledge that is fundamental to the use of a skill. An example may consist of a programming task. In order to effectively write a script one needs good logical and analytical competences as well as the skill to write a program in a specific language. Hence, Java is a skill. Although, underlying the ability to use that skill effectively is a competence.

2.1.2 Metrics between partial rankings

A ranking is a one way to compare the performance of individuals and teams. We use this notion in the experiments done in chapter 4 to compare teams' performance and in Chapter 6 to produce a ranking based on the aggregated opinions of reviewers. Notice that these rankings may include ties since several objects may be equally valued. An ordering with ties is also known as a *partial ranking*. Given two different aggregation methods for producing partial rankings, we are interested in comparing them to decide which aggregation method is better. For this purpose, we require metrics to compare partial rankings. The purpose of this subsection is to introduce such metrics. With this aim we largely rely on the work by Fagin el tal. [Fagin et al., 2004][1] , which provides sound mathematical principles to compare partial rankings. In particular, we will detail one of the four metrics presented in [Fagin et al., 2004], the so-called *Kendall distance with penalty parameter p*. Before that, we require some preliminary definitions.

Definition 2.1 (Bucket order)**.** A bucket order is, intuitively, a linear order with ties. Formally, given a domain D, a bucket order is a transitive binary relation $\lhd$ for which there are sets $\mathcal{B}_1, \cdots, \mathcal{B}_t$ (the buckets) that form a partition of D such that $x \lhd y$ if and only if there are i, j with $i < j$ such that $x \in \mathcal{B}_i$ and $y \in \mathcal{B}_j$.

[1]We refer the reader to [Fagin et al., 2006] for a more detailed, extended version on the topic.

A bucket contains objects that are "tied". We say that $\mathcal{B}_i$ is the bucket of x if $x \in \mathcal{B}_i$. We say that bucket $\mathcal{B}_i$ *precedes* bucket $\mathcal{B}_j$ if $i < j$. Thus, $x \lhd y$ if and only if the bucket of x precedes the bucket of y.

Notice that a linear order is a bucket order where every bucket is of size 1.

Definition 2.2 (Bucket position)**.** Let $\mathcal{B}_1, \cdots, \mathcal{B}_t$ be a bucket order over D. The position of a bucket $\mathcal{B}_i$ in the bucket order is defined as $pos(\mathcal{B}_i) = (\sum_{j<i} |\mathcal{B}_j|) + (|\mathcal{B}_i| + 1)/2$.

Intuitively, $pos(\mathcal{B}_i)$ stands for the average location within bucket $\mathcal{B}_i$. Now, we can formally define the notion of partial ranking based on the notion of bucket order.

Definition 2.3 (Partial ranking)**.** Given a domain D and a bucket order $\mathcal{B}_1, \cdots, \mathcal{B}_t$ over D, the partial ranking σ associated with the bucket order is a function that maps each element in D to the position of its bucket, namely $\sigma(x) = pos(\mathcal{B})$ when $x \in \mathcal{B}$.

Given a partial ranking σ, we say that *x is ahead of y in σ* if $\sigma(x) < \sigma(y)$, and that *x and y are tied in σ* if $\sigma(x) = \sigma(y)$.

Now, let $\mathcal{P} = \{\{i,j\} | i \neq j \ \ and \ \ i,j \in D\}$ be the set of all the unordered pairs of different elements in D. Given two partial rankings σ_1 and σ_2 with domain D, we will define a penalty measure $\bar{K}^{(p)}_{i,j}(\sigma_1, \sigma_2)$ to account for the different ordering of i, j in partial rankings σ_1,σ_2, where p is a fixed parameter such that $0 \le p \le 1$. We shall distinguish three cases:

Case 1: i and j are in different buckets in both σ_1 and σ_2. (i) If i and j are in the same order in σ_1 and σ_2 (e.g. $\sigma_1(i) > \sigma_1(j)$ and $\sigma_2(i) > \sigma_2(j)$) then $\bar{K}^{(p)}_{i,j}(\sigma_1, \sigma_2) = 0$, and thus there is no penalty for $\{i, j\}$. (ii) If i and j are in the opposite order in σ_1 and σ_2 (e.g. $\sigma_1(i) > \sigma_1(j)$ and $\sigma_2(i) < \sigma_2(j)$) then let the penalty $\bar{K}^{(p)}_{i,j}(\sigma_1, \sigma_2) = 1$.

Case 2: i and j are in the same bucket in both σ_1 and σ_2. Since both partial rankings agree that i and j are tied, there is no penalty and $\bar{K}^{(p)}_{i,j}(\sigma_1, \sigma_2) = 0$

Case 3: i and j are in different buckets in only one of the partial rankings. In this case, the penalty is $\bar{K}^{(p)}_{i,j}(\sigma_1, \sigma_2) = p$.

Now we are ready to define the Kendall distance between two partial rankings.

Definition 2.4 (Kendall distance)**.** Given two partial rankings σ_1 and σ_2 over domain D, we define their $K^{(p)}$, their Kendall distance with parameter p, as follows:

$$K^{(p)}(\sigma_1, \sigma_2) = \sum_{\{i,j\} \in \mathcal{P}} \bar{K}^{(p)}_{i,j}(\sigma_1, \sigma_2).$$

Notice that from the definition above, we can readily define a normalised version of the Kendall distance that we will employ in this paper to compare partial rankings.

Definition 2.5 (Normalised Kendall distance)**.** Given two partial rankings σ_1 and σ_2 over domain D, their normalised Kendall distance with parameter p is defined as:

$$\tilde{K}^{(p)}(\sigma_1, \sigma_2) = \frac{K^{(p)}(\sigma_1, \sigma_2)}{s}$$

where $s = \frac{|\mathcal{P}|\cdot(|\mathcal{P}|-1)}{2}$ is the number of pairs in $\mathcal{P}$.

Finally, notice that the work in [Fagin et al., 2004] defines three further metrics to compare partial rankings, which also admit efficient computation. However, it does not matter the metric that we choose because the equivalence results in [Fagin et al., 2004] indicate that the four metrics are all within constant multiple of each other.

2.2 Individual Assessments

Having introduced the background for the thesis, we can move on and discuss the literature related to this work. We start from categorising the most recent literature related to peer assessments.

Artificial Intelligence research has focused on the assessment process for long and a number of algorithms have been developed to assist in assessing the performance of humans or artificial agents. Indeed, large number of trust and reputation models have been proposed [Alfaro and Shavlovsky, 2013; Lu and Zhang, 2012; Osman et al., 2015, 2010; Piech et al., 2013; Stepanyan et al., 2009; Topping, 1998; Walsh, 2014; Wu et al., 2015a; Zhang et al., 2007]. In this subsection we discuss the main research approaches dealing with quantitative analysis of peer review.

Table 2.1 categorises the related work with respect to whether they weigh assessments by their reliability (where WbR stand for 'Weighed by Reliability'). We discuss these models next.

Table 2.1: Categorisation of Individual Assessment Models

WbR	¬ WbR
CrowdGrader [Alfaro and Shavlovsky, 2013], PeerRank [Walsh, 2014], Piech et al. [Piech et al., 2013], [Wu et al., 2015b]	LocPat [Hang and Singh, 2012], Collaborative Filtering [Shardanand and Maes, 1995], Simple aggregation (mean or median)

2.2.1 Models weighted by Reliability

CrowdGrader [Alfaro and Shavlovsky, 2013] is a framework which defines a crowdsourcing algorithm for peer assessments. The authors claim that, when performing assessments, relying on a single person is often impractical and can be perceived as unfair. Their method aggregates the assessments of an assignment made by several students into an overall assessment for the assignment, relying on a reputation system. The reputation of each student (or their *accuracy degree* as they call it) is measured by comparing the student's assessments with the assessments of their fellow students for the same assignments. In other words, the reputation of a student describes how far are her assessments from those of her fellow students. The overall assessment (consensus grade) is calculated by aggregating all student assessments weighted by the reputation of the students providing them. The algorithm executes a fixed number of iterations using the consensus grade to estimate the reputation (or accuracy degree) of students, and then uses the updated student's reputation to compute more precise suggested assessments. However, one's assessment does not need be similar to others, but needs to be highly viewed by others. For instance, think of the clever student who always makes excellent observations that have gone unnoticed by others.

PeerRank [Walsh, 2014] is based on the idea that the grade of an agent is constructed from the grades it receives from other agents, and the grade an agent gives to another agent is weighted by the grading agent's own grade. Thus, the grade of each agent α is calculated as a weighted average of the grades of the agents evaluating α, and thus the grades of α's evaluators are themselves weighted averages of the grades of other agents evaluating them, and so on. The final grades are defined as a fixed point of an equation, similar to PageRank, where web-pages are ranked according to the ranks of the web-pages that link to them.

Piech et al. [Piech et al., 2013] propose a method to estimate student reliability and to correct student biases in an online learning scenario, presenting results over two Coursera courses. They assume the existence of a true score for every assignment, which is unobserved and to be estimated. Every grader is associated with a bias, which reflects the grader's tendency to inflate or deflate her assessments with respect to the true score. Also, graders are associated with a reliability which reflects how close the grader's assessments tend to land near the corresponding true score, after having them corrected for bias. Authors infer the values of these unobserved variables using known approximated inference methods such as Gibbs sampling. The model proposed is therefore probabilistic and is compared to the grade estimation algorithm used on Coursera's platform (mean of assessments), which does not take into account individual biases and reliability.

Wu et al. [Wu et al., 2015b] investigate consensus building between a group of experts in a trust network. New trust relationships are derived from the trust network and the trust scores of such relationships are calculated using an averaging operator that aggregates trust/distrust values from multiple trust paths in the network. The trust score is used to distinguish the most trusted expert from the group and, ultimately, to drive the aggregation of the individual opinions in order to arrive at a group consensual decision making solution. This work also includes a visual consensus model to identify discordant opinions, to produce recommendations to those experts that are furthest from the group, and to show future consensus status if experts are to follow the recommendations.

2.2.2 Models not weighted by Reliability

The important group of models that do not use reliability to calculate final opinion are recommender systems. Recommender systems tune their results to the point of view of a specific person. An interesting example can be system LocPat [Hang and Singh, 2012] that is a generalised framework for personalised recommendations in agent networks. LocPat builds trust measures based on mining the graph of an agent network. For instance, trustworthy relationships are discovered by studying the link structure (e.g., the number of common neighbours). Then, it suggests to a specific requester (who requests a recommendation in the agent network) a list of trustworthy agents for the requester to interact with.

Collaborative Filtering [Shardanand and Maes, 1995] is a classical social information filtering algorithm that recommends content to users based on their previous ratings, exploiting similarities between the tastes of different users. In summary:

1. The system maintains a user profile, which is a record of the user ratings over specific items.
2. Then, the system computes a similarity measure among users' profiles.
3. Finally, the system recommends items to users with a rating that is a weighted average of the ratings on that item given by other users. The weights are the similarity measures between the profiles of users rating the item and the profile of the user receiving the recommendation.

In sext section we discuss the literature relevant to team aspects of this thesis.

2.3 Team and Organisation literature

In this section we review the contributions in both the computer science literature and the organisational psychology literature dealing with topics of Team Composition and Formation. Our purpose is twofold. First, we aim at identifying the strengths and weaknesses of the contributions made by these two diverse bodies of research. Second, we pursue to identify cross-fertilisation opportunities that help both disciplines benefit from one another.

Team research in MAS has considered a variety of application domains (e.g. Unmanned Aerial Vehicle (UAV) operations [Haque et al., 2013], teamwork in social networks [Lappas et al., 2009] or RoboCup rescue teams [Ramchurn et al., 2010]) wherein agents face the challenge of performing tasks that are either too complex for one single agent or limited in time, thus requiring several agents to collaborate.

Nevertheless, research on team composition and team formation in computer science (CS) and organisational psychology (OP) has evolved separately. On the one hand, MAS literature has typically disregarded significant OP findings, with the exception of several recent, preliminary attempts (such as [Farhangian et al., 2015a; Hanna and Richards, 2015]). Thus, this body of research has focused on algorithms that help automate team formation and composition. On the other hand, the OP literature has mainly focused on empirically investigating the factors that influence team performance to develop heuristics that help organisations handcraft their teams. OP has disregarded the algorithmic results developed by computer scientists to automate team composition and formation. Despite the common research interests shared by MAS and OP, to the best of our knowledge there has been no effort in the literature to bridge the knowledge produced by both research disciplines.

Against this background, we would like to survey both disciplines, to analyse and compare the strengths and weaknesses of their contributions, and to identify research gaps and opportunities by bringing together the knowledge of the two research strands on team composition and formation. This analysis also pursues to identify cross-fertilisation opportunities that help both disciplines benefit from one another.

In order to structure our analysis, we have identified several dimensions that help us dissect the contributions from both research fields:

1. *WHO is concerned?* The attributes of the agents involved.
2. *WHAT is the problem?* The features of the task to complete by a team.
3. *WHY do we do it?* The objective function to optimise when composing/forming a team.
4. *HOW do we do it?* The organisation and/or coordination structure adopted by the team in charge of performing a particular task.
5. *WHEN do we do it?* The dynamics of the stream of tasks to be completed by agent teams.

6. *WHERE do we do it?* The context wherein team composition/formation occurs.

Our analysis of the literature indicates that Computer Science (CS) and Organisational Psychology (OP) exhibit some similarities. Indeed, one of the crucial findings in both OP and CS is that team members have to be heterogeneous to maximize team performance. When modeling agents, CS and OP agree on considering two main approaches: either there is complete information about the attributes of each agent; or agents are capable of learning about their teammates through repeated interactions. Regarding tasks, both OP and CS research largely focus on finding team members whose attributes make them capable of performing a given task based on its requirements. In other words, they are both concerned with matching agents (or whole teams) with tasks.

However, there are important differences between the contributions made by OP and CS that stem from the fact that OP does consider the whole complexity of: humans as team members, tasks, the *context* where teams perform tasks (understood as the internal and external factors influencing teamwork), and the dynamics of the actual-world scenarios where tasks appear to be serviced. Thus, OP assumes that human capabilities are necessarily dynamic (evolve along time) so that teams can successfully perform tasks in dynamic real-world scenarios and in a variety of contexts. Furthermore, OP observes that the quality of human resources (e.g. motivation, satisfaction, commitment), the ability of individuals to learn new capabilities, and the context constraining team performance significantly influence team performance. Finally, OP research also focused on identifying correlations between task types and team types to compose the best team depending on the type of each particular task. All these findings contributed by OP research offer interesting opportunities for cross-fertilisation.

The rest of this chapter is organised as follows. Section 2.1 introduces some fundamental terminology to make clear what we mean by team composition, team formation and teamwork. Thereafter, the remaining of this chapter is organised around two main sections. Section 2.4 reviews the MAS contributions to team composition and team formation. Next, section 2.5 surveys the contributions in the organisational psychology literature. Finally, section 2.6 identifies the main similarities and differences between the two bodies of research. Furthermore, it also discusses cross-fertilisation opportunities between both fields that may spur future research.

2.4 Team engineering in Computer Science

Team composition and formation are critical issues for co-operative multiagent systems. In this section we survey the most recent and representative approaches in the MAS literature to the team composition and formation problems along the dimensions identified in the introduction above.

2.4.1 WHO is concerned?

The question behind team composition and formation is how to create a multiagent system as a group of heterogeneous agents (such as humans, robots, software agents or even animals) and how to organise their activities. Team members must observe the environment and interact with one another in order to perform tasks or solve problems that are beyond their individual capabilities. The algorithms to create these teams take inspiration from human teamwork. We observe people working together on daily activities as well as on research and business projects. For instance, there are sport teams (e.g. football, basketball), police squads, search and rescue teams formed by dogs and humans, and we start to witness human-robot cooperation in houses, hospitals, or even in space missions [Hoffman and Breazeal, 2004].

In general, MAS research focuses on the interaction among intelligent agents. In the team formation literature, the focus is on the interaction of cooperative and heterogeneous agents. That is, agents who share a common goal, and have different individual attributes. Therefore, in this section, we would like to account for the different ways previous research has dealt with these questions. We will classify individual attributes according to two dimensions:

1. Capacity: individual and social capabilities of agents; and

2. Personality: individual behaviour models.

2.4.1.1 Capacity: individual and social capabilities of agents

In many domains, a capability is defined as a particular skill required to perform an action. The capacity dimension has been exploited by numerous previous works, like Robust Team Formation [Crawford et al., 2016; Okimoto et al., 2015] or Online Team Formation [Anagnostopoulos et al., 2012]. In these works, agents are assumed to have multiple binary skills (i.e., the agent either has a required skill or not). This is a simplistic way to model an agent's capabilities since it ignores any skill degree. In real life, capabilities are not binary since every individual (e.g. human or robot) shows different action performance. This is why some works propose a more realistic approach by defining graded agent capabilities, for instance by defining skill levels [Chalkiadakis and Boutilier, 2012].

On a different vein, [Rangapuram et al., 2013] builds a weighted, undirected graph where the weight between each pair of agents reflects their degree of compatibility to jointly solve tasks. These weights are updated along multiple encounters between agents. In a somehow related vein, [Peleteiro et al., 2015] try to capture the quality of the solutions of team tasks via a model that besides using skills and compatibility between agents (called the strength of collaboration synergies within coalitions), calculates the reputation of teams (coalitions) as a whole and of single agents. These reputation values are used by the team composition process.

Typically, the capabilities of agents are assumed to be known, though there exist models that consider that an agent can learn the capability levels of other agents. For instance, [Liemhetcharat and Veloso, 2014] had the insight that repeated interactions allow to discover the capabilities of other agents. They call "synergy" to the degree of performance of a team. Agents learn a model of synergy via repeated interactions. Such synergy values are then used by individual agents to learn the capabilities of others, and hence to subsequently compose teams with improved performance. However, in open environments (that is, when new agents and tasks are dynamically introduced), agents need more sophisticated procedures to decide which team to join. For instance, [Chen et al., 2015] propose an ad-hoc team formation framework that considers learning other agents' capabilities in the context of unknown tasks. In order to solve a new task, agents would prefer to team up with unknown agents instead of with agents whose known capabilities do not adjust to the task. They observe that learning the capabilities of others in the context of agent and task openness improves team composition and task resolution.

2.4.1.2 Personality: Individual behaviour models

Personality is key to understand people's behaviour, cognition and emotion. The use of personality models in agents helps to create more realistic complex scenarios. Indeed, autonomy is related to how individuals behave and what makes them behave differently, even when facing the very same situation. Personality provides a mechanism for behaviour selection that is independent of social background (such as beliefs or morality). Very recently some MAS contributions have started to consider the notion of personality, i.e. individual behaviour model, to compose heterogeneous teams. For instance, [Hanna and Richards, 2015] study the influence of two agent personality traits: extraversion and agreeableness, both expressed as verbal and non-verbal communication skills. They construct pairs of human users and Intelligent Virtual Agents (IVAs) and analyse how the personality traits influence the development and maintenance of a Shared Mental Model (SMM). The results confirm the importance of providing IVAs with these personality traits to succeed in jointly solving tasks.

Marcolino et al. [Marcolino et al., 2013, 2016; Nagarajan et al., 2015] propose a new approach for action selection. A task is a sequence of actions to be decided at execution time. To choose which action to execute next, every heterogeneous agent within a team votes for its preferred candidate action. Agents vote according to a probability distribution over actions that varies for each agent. This can be understood as a way of modeling an agent's personality, motivations and beliefs (causing him to behave in a certain way).

In a series of papers, [Farhangian et al., 2015a,b] use the Myers-Briggs Type Indicator (MBTI) [Myers et al., 1998] scheme to model different agent personality types. [Farhangian et al., 2015b] use both individuals' skills and personality types (measured by MBTI and Belbin [Belbin, 1993] personality tests) to compose teams. These two dimensions are used to simulate human team composition in a business environment.

Another aspect covered by the existing literature is the individual agent knowledge about the other team members' personalities, that is, about their behaviour models. These works go beyond many "ad-hoc" team composition systems where information details about the behaviour of individual agents is absent. [Barrett et al., 2013] focus on how a new member in a team behaves in order to cooperate well with the other team members whose behaviors are unknown. Each agent is endowed with a learning mechanism for building models of the behaviours of many distinct types of other agents via repeated interactions. A similar setting is presented by [Agmon et al., 2014], though they consider that there are only two types of agents: a best response agent (choosing his action based on the current state of the world), and an ad-hoc agent (has a better awareness of the team's possible actions and the resulting joint utility). There is no a-priori model, hence, similarly to [Barrett et al., 2013], an ad-hoc agent needs to decide his behaviour by observing his peers.

Analysis. In summary, team composition and formation research has focused so far on cooperative, heterogeneous agents that have a set of attributes. These attributes can be categorized into two groups: capacity and personality. To our knowledge, besides [Farhangian et al., 2015b], there has been no further attempts to combine capabilities and personality for team composition and formation in the area of MAS. Besides that, we observe that the capabilities of agents are always static, but the behaviour model may change with agents' interactions. While the capabilities of humans change over time, the MAS literature typically does not consider dynamic capabilities for software agents. Finally, when modeling agents' attributes, many existing approaches typically assume extensive a-priori information about teammates. This is a strong limitation for real-life settings. Notice that in many companies there is no central and extensive knowledge about all employees' capabilities.

2.4.2 WHAT is the problem? The notion of task

In its most general sense, a task is a course of action to achieve a goal. The execution of a task is then usually equated to the execution of an action plan. Action plans can be rather complex as they may take into account concurrency of actions, time constraints, action order, or environment uncertainty. However, in the team formation literature it is often the case that simplifying assumptions are made and tasks are assumed to be solved by simple action plans. For instance, an action plan can be seen as a set of actions, or even as a set of competences. In this latter case the idea behind is that the task can be successfully solved by a team of individuals with expertise in a number of different fields. In this section, we review which concepts of task have been proposed in team formation and team composition. We identify two main approaches:

- Individual-based, i.e. capacity or personality (see section 2.4.1);
- Plan-based, e.g. the set of actions or subtasks.

Next we discuss each approach in detail.

2.4.2.1 Individual-based approaches

Sometimes teams work less effectively than initially expected due to several reasons: a bad balance of their capacities, bad personal relations, or difficult social situations. Hence, in order to make sure a task is performed the most effectively, the large body of literature defines the action plan of the task as a set of requirements for agent individual characteristics. It is assumed that the task can be fulfilled if the task requirements are a subset of the capabilities of team members. We categorise existing work on team composition with the purpose to solve a task into two categories of individual attributes: capacity and personality.

Capacity. The capabilities of team members are crucial while performing a task. For instance, it is obvious that in order to develop an online Java application, the collective team knowledge has to include Java, Java EE, front-end tools, and database and server knowledge. In the MAS literature (as discussed in Subsection 2.4.1.1), the majority of research work expresses capabilities as binary (they are present or they are not) [Anagnostopoulos et al., 2012; Chen et al., 2015; Crawford et al., 2016; Okimoto et al., 2015]. The main shortcoming of the binary approach is the restrictive assumption that if an agent has a capability, his expertise level is sufficient to perform a given task, which implies that the quality of the task performed is not relevant.

In many cases, the definition of a task is indirectly connected to the agents' capabilities. [Peleteiro et al., 2015] propose a model where a task is defined as a tuple that contains the specification of the task (i.e. its subtasks) and the deadline by which the task has to be completed. Each subtask is then matched with one capability. A contract net algorithm is used to compose a team of agents that covers all the required capabilities while maximizing the reputation of the team, thus leading to the best expected performance. In [Chalkiadakis and Boutilier, 2012], a *project* is defined as a set of tasks, where each task has a complexity level (e.g. moderate or ambitious). Agents' capabilities are graded (e.g. a good carpenter). Tasks are matched with agents' capabilities. The probability of an agent succeeding at performing a task depends on the capability degree of the agent performing the task and the complexity level of the task. These probabilities are learned through repeated interactions between agents, and then used by them to self-organise as teams. Finally, in Roles and Teams Hedonic Games (RTHG) [Spradling et al., 2013] each agent expresses his preferences over both his own roles within a team and on the set of roles needed in the team. This way, agents themselves jointly select a set of required capabilities to perform a given task.

Personality In [Farhangian et al., 2015a], the nature (structure) of a task is quantitatively characterized: from extremely structured to extremely open-ended. While structured tasks are straightforward and do not require planning, open-ended tasks require creativity and imagination from team members. In another article, [Farhangian et al., 2015b] try to capture the dynamics of tasks by matching the required levels of creativity, urgency, social interaction and complexity of a task to personalities of agents. For instance, teams composed of differing attitude tendencies (associated with different personalities) are believed to outperform teams composed of like-minded people when tackling tasks requiring a high level of creativity.

[Hanna and Richards, 2015] show that when performing a task, the personality of team members influences their success. They analyse the influence of an Intelligent Virtual Agent (IVA) communication style (expressing its personality) on human-IVA cooperation. The task is a collaborative game that involves dodging a sequence of obstacles to reach a target.

2.4.2.2 Plan-based approaches

The notion of task in plan-based approaches is normally understood either as a set of actions or as a sequence of actions. Well organised teamwork can shorten the time required for completing a particular task by distributing a set of actions across team members. Both [Barrett et al., 2013] and [Agmon et al., 2014] employ an indirect planning method driven by the most informed agents to solve a set of actions. [Barrett et al., 2013] introduce an ad-hoc team agent that learns its teammates' models (i.e. their predictable action selection) and chooses its own actions so that they collectively maximize the likelihood of success. In detail, they use Monte Carlo sampling to simulate the long term effects of collective actions. As an extension to the previous work, in [Agmon et al., 2014] the actions selected by ad-hoc agents influence the actions that the other team members will choose. Each agent has a set of possible actions that it may choose in order to solve each subtask. The ad-hoc agents need to predict the actions of its teammates (conditioned in this case to its own actions) and behave based on these predictions with the purpose of influencing the collective selection of actions in the team to reach a joint optimal solution.

Among the approaches considering a task as a sequence of actions, in [Marcolino et al., 2013] a team of agents jointly playing the computer game Go plan which action to take next by voting on the possible alternatives from a discrete set of possible actions. Authors prove that under certain conditions of opinion diversity, aggregating the decisions of a team of heterogeneous agents is a better planning strategy than the decision of a team built with copies of the most competent agent (called the strongest agent). This shows that diversity improves the planning capacity of a team solving a complex task like Go. In [Marcolino et al., 2016], the authors use the same technique to suggest a user a number of optimal solutions for their next action decision. The application domain of their algorithm is house design. Various design alternatives are proposed to the user in order to select one for further study.

Similarly, in [Moon et al., 2005] the plan is created by team members during a game. The domain used for this study is an on-line multi-player computer game called America's Army, which is a first-person shooting (FPS) game. The game is the duel of two teams, usually an assault team and a defense team. A team consists of one to fourteen players. Every game starts with a new set of players that need to coordinate their activities during the game in order to win. Players are allowed to communicate in a team chat. A team wins the game either by killing all of its opposing players, or by accomplishing the goal for that mission (for instance, securing an oil pipeline or crossing a bridge).

Finally, [Rochlin et al., 2016] deal with self-interested agents in a team that select one agent to accomplish the task of purchasing a jointly desired item with the lowest possible cost. By doing so, the team assigns the execution of the plan to a single member of the team, becoming the buyer. The buyer's strategy decides whether to maintain the search looking for better deals (search for a further action), or stop looking and buy at the lowest price found so far, bearing the incurred buyer's overhead. This strategy balances the expectation of finding a better price (considering the price distribution built during the search) and the team policy to reimburse the cost of the task solution finding to the buyer.

Analysis. In conclusion, tasks are solved by the execution of action plans. How complex these action plans are depends on the focus of the reviewed contributions. Individual-based approaches understand action plans as sets of requirements on a team members' capacity and personality. These approaches assume that the joint capabilities of agents in a team must be enough to solve a given task. Contrarily, plan-based approaches regard tasks as sets of actions or sequences of actions that are assigned to the individual members of a team. All these works propose algorithms that determine which action will be executed and by whom. However, plan-based approaches have a very simplistic notion of plan. The majority of models do not consider time constraints, action dependencies, action failure, plan robustness, or dynamic changes in a task requirements. Therefore, the vast literature on planning has not yet been integrated into team formation methods.

2.4.3 WHY do we do it? The objective(s)

The motivation of individual efforts or actions is to attain or accomplish a certain state of affairs: its goal. A necessary condition for a team to exist is that all team members are committed to a joint goal. Therefore, in Computer Science an agent team is typically built of at least two cooperative agents that share a common goal; by teaming up, these goals can be achieved in a more effective way. This is the main motivation of team composition and formation. A large body of literature proposes team composition algorithms to attain at least one of the following team objectives:

1. minimizing overall cost (e.g. cooperation cost, team cost);
2. maximizing social utility; or
3. maximizing the quality of an outcome.

In this section we describe the literature on team composition per objective.

2.4.3.1 Minimizing overall cost

Team cost efficiency has received some attention in the literature. There are various costs associated with team composition and formation problems (e.g. communication costs or agent service costs). For instance, some results balancing cost and quality were obtained by [Kargar et al., 2012]. They propose algorithms for composing a competent team in a social network. When composing a team, those algorithms minimize team members' costs and communication costs within the team. [Kargar et al., 2012] require that agents have the necessary competences to perform a task, but do not require any specific motivation from them.

A similar approach is presented in [Crawford et al., 2016] and [Okimoto et al., 2015]. These works propose a model for robust team composition and go a step further with respect to [Kargar et al., 2012] since they minimize the overall cost among k-robust teams (see Section 2.4.4.1 for a definition of a k-robust team). That is, this model assumes that up to k agents within a team may eventually fail without affecting the achievement of the task. Thus, it assumes more realistic conditions than [Kargar et al., 2012]. However, likewise [Kargar et al., 2012], agents' motivations to work together in a team are not considered. Finally, [Anagnostopoulos et al., 2012] propose approximation algorithms to compose teams minimizing simultaneously low coordination costs and agent workload.

2.4.3.2 Maximizing social welfare

A second objective considered in the team composition and formation literature is maximizing social welfare. That is, maximizing the utility function of a team, as a whole, while performing a task. The utility obtained is then allocated to the individual members of the team. For instance, [Chalkiadakis and Boutilier, 2012] propose a Bayesian Reinforcement Learning framework where agents learn from iterated coalition compositions. Agents can choose between exploration (select coalitions to learn more about new agent types) and exploitation (rely on known agents). Exploitation enables agents to maximize their utility function by performing tasks with reliable agents (discovered during the exploration phase).

Paradoxically, the agent motivation to maximize its individual welfare may reduce the overall team cost and additionally increase the overall quality of the performed task. For instance, in [Rokicki et al., 2015] a human team competition mechanism improves cost efficiency and the quality of a solution in a team-based crowdsourcing scenario. In conventional crowdsourcing reward schemes, the payment of online workers is proportional to the number of accomplished tasks (pay-per-task). Rokicki et al. examine the possibility of getting much higher rewards by introducing strategies (e.g. random or self-organised) for team composition. Their mechanism triggers the competition among human teams as the reward is only given to the top-5 performing teams or individuals. Their evaluation shows substantial performance boosts (30% in the best scenario) for team-based settings without decreasing the quality of the outcome.

The objective of maximizing social welfare is also considered in many ad-hoc settings, like the one proposed by [Agmon et al., 2014]. Agmon et al. consider a framework with two types of agents: best-response and ad-hoc agents forming teams. On the one hand, best-response agents have limited knowledge and assume that the environment and their teammates will behave as observed in the past. On the other hand, ad-hoc agents have a more complete view of a team actions, agents' joint utilities and their action costs. Using such information, ad-hoc agents try to influence joint decisions. In [Agmon et al., 2014] the authors consider that ad-hoc agents know with uncertainty their teammates' behaviour. The paper analyses the impact on optimal solutions of ad-hoc agents misidentifying their teammates' types.

The study of self-interested agents that co-operate in a team has also attracted the interest of researchers in MAS. An interesting example of this approach is presented in [Farhangian et al., 2015a], where self-interested agents need to maximize the welfare of all team members in order to maximize their own benefit. Hence, they indirectly aim at maximizing the utility of the team. Similarly, in [Chen et al., 2015] agents repetitively decide which team to join by balancing both rewards from completing tasks and learning opportunities from more qualified agents. That is, each agent consider whether to sacrifice short-term rewards to acquire new knowledge that benefits himself and the whole community in the long run.

2.4.3.3 Maximizing quality

The last range of models propose a number of methods where agents try to maximize the quality of solutions whilst minimising the time to achieve them, namely to maximize team performance.

Recent organisational psychology studies show that team members' diversity is a key factor to increase team performance [Wilde, 2009]. As mentioned in Section 2.4.1 [Marcolino et al., 2013] present a setting where agents in a team vote together to decide on the next joint action to execute that maximises the team's solution quality. The authors show that a diverse team can overcome a stronger team (i.e. a team built of copies of the strongest agent) if at least one agent has a higher probability of taking the best action in at least one world state than the probability that the best agent has of taking that action in that state.

[Hanna and Richards, 2015] also use personality to investigate the influence of Intelligent Virtual Agents (IVA) on team collaboration. Their findings reveal that team performance boosts when the human and the IVA in a team have a shared mental model. Building a shared mental model is directly related to the psychological traits of IVA.

[Carley et al., 2005] found that the most favorable size of a team is ten because of the relatively higher survival ratio. Also, frequent usage of the weapon, precision of the weapon used, and frequency of communication, can be the distinctions between winning teams and losing teams. Moreover, frequent communication increases a team's situation awareness, that is, gives information about where other team members are and how they can be supported.

[Peleteiro et al., 2015] introduce a decision making mechanism that on top of improving the quality, aims at increasing the quantity of completed tasks. It uses reputation and adaptation mechanisms to allow agents in a competitive environment to autonomously join and preserve coalitions (teams). In terms of team performance, they show that coalitions keep a high percentage of tasks serviced on time despite a high percentage of unreliable workers. Moreover, coalitions and agents demonstrate that they successfully adapt to a varying distribution of incoming tasks.

[Liemhetcharat and Veloso, 2012] developed a model to learn and analyze capabilities of agents and synergies among them to solve the team composition problem using previous joint experiences. They define a synergy model as a graph where the distance between agents is an indicator of how well they work together. Their main contribution is that their algorithm learns from only a partial set of agent interactions in order to learn the complete synergy model. In a subsequent article [Liemhetcharat and Veloso, 2014], the authors study the learning agent team formation problem with the goal of maximizing the mean performance of a team after K learning instances. There, learning agent pairs have heterogeneous rates of coordination improvement, and hence the allocation of training instances has a larger impact on the performance of the final team.

The notion of fairness is also considered in the context of team performance. An example of this approach is given in [Rochlin et al., 2016]. Rochlin et al. analyze the correlation between efficiency and fairness in teams consisting of self-interested agents. They prove that the more fair the team the more efficient its members are.

Finally, it is worth discussing how researchers in computer science evaluate and monitor the achievement of the objectives mentioned above. Omitting this information can create a false equivalence between the findings of research studies conducted in very different conditions. We distinguish among three main data sources, that is:

- Existing databases available online containing real data,
- Data simulation,
- Empirical data.

Existing data Finding ready datasets for validation of team composition and formation problems is challenging. Systems supporting team composition or/and formation are not yet in broad use and most data from them is not publicly available. For this reason, some authors use bibliography (such as Citeseer, DBLP, Bibsonomy), movie datasets (IMDb) or a software engineering environment (such as the Python Enhancement Proposals (PEP)) that can demonstrate the effectiveness of their approach [Anagnostopoulos et al., 2012; Farhangian et al., 2015a; Kargar et al., 2012; Rangapuram et al., 2013]. For instance, [Rangapuram et al., 2013] use an academic scenario (Citeseer database) to perform a qualitative and quantitative assessment of teams.

Data simulation The most common approach to test team composition and formation algorithms is to perform a set of simulations showing the effectiveness of team methods. The majority of researchers simulating data use an abstract set of simulated tasks. Depending on the model, tasks can be static [Crawford et al., 2016; Liemhetcharat and Veloso, 2012; Okimoto et al., 2015; Peleteiro et al., 2015] or dynamic, that is, they can change over time [Chalkiadakis and Boutilier, 2012; Farhangian et al., 2015b]. [Chen et al., 2015] use both static and dynamic tasks to study various effects of considering agent openness (AO) and task openness (TO) in ad-hoc team formation.

Empirical data Collecting empirical data is time consuming, however it is the most reliable way to validate team hypotheses and models. The data can be collected in real world (mostly robotics) or in virtual environments (such as on-line games). For many years, RoboCup has served as an excellent domain for testing teamwork, coordination, and cooperation. In 2013, a new competition began that serves as a testing environment for cooperation without pre-coordination: The Drop-in Player Competition. In this competition, instead of homogeneous teams of robots such as all robots are programmed to follow the same strategy, all robots are heterogeneous (originating from different RoboCup teams and as such running different software). [Genter et al., 2016] present their findings from a three year experiment in the domain above that consisted of 38 games for a total playing time of 510 minutes that resulted in approximately 85 robot hours. The authors suggest improvements to the competition, and provide advice for organising new competitive ad-hoc teamwork evaluations.

An example of teamwork conducted in a virtual environment is [Hanna and Richards, 2015], which uses human-agent teams to assess the performance of teams. The results show the importance of designing agents capable of using multiple methods of communication with humans, as this tends to build shared mental models with human users and improve team performance. [Rokicki et al., 2015] use a crowdsourcing scenario for a face recognition task where human agents (workers) are asked to identify a person on a given reference photo among a set of 10 test photos. The performance is evaluated by the quality of the final outcome of each team. Many researchers use on-line games to do team performance studies. For instance, [Marcolino et al., 2013] validate their hypothesis by using virtual agents playing 691 instances of the GO computer game, and [Moon et al., 2005] analyse the behaviour of approximately 150.000 teams in America's Army game. Finally, [Andrejczuk et al., 2017] use an education scenario to pitch their automated team composition model with the team composition performed by experts. Authors compare both team composition models in terms of how well they predict team performance.

Analysis. In summary, the computer science literature has focused on team co-operation with various objectives that can be categorized as at least one of the following: minimizing overall cost, maximizing social utility, or maximizing team(s) performance. The models minimizing overall cost compose teams based on individual competences, though do not take into account individual motivations to complete the assigned task. This is a rather strong assumption, especially when it comes to mixed teams or human teams, making the existing approaches rather unrealistic. The literature focusing on maximizing social welfare considers both agent competences and motivation. The motivation increases by using competence mechanisms (like in crowdsourcing teams), or by giving agents the freedom to select their collaborators (like in learning agent team formation or in ad-hoc teams). To maximise team performance, one of the crucial findings in both Organisational Psychology and Computer Science is that team members must be heterogeneous. Further variables that have been used by computer scientists in the area of MAS to compose teams are: agent reputation, personality of humans and agents, synergy between team members, and feeling of fairness among team members. The proposed methods are validated using existing databases available online, data simulation or empirical data.

2.4.4 HOW do we do it? The organisation

In the existing literature, the societal structure of teams is considered crucial for effective teamwork. There are two aspects to be considered, one is which agents will be members of a team and second, how teams will be organised to solve tasks. Thus, the different approaches in the literature can be classified depending on the functionality that they tackle:

- Team Composition: the process of deciding which agents will be part of a team. It can be an external decision or an autonomous decision by the agents themselves; and
- Team Formation: the process of learning to decide the roles and internal organisation of a team. This organisation can be imposed or be the result of self organisation. In any case, the resulting organisations can be categorized as hierarchical or egalitarian.

Next, we look into these two dimensions in detail.

2.4.4.1 Team Composition.

Although team composition in MAS has mainly focused on building teams of software agents, that is agent teams, there is a growing number of works considering either mixed teams [Hanna and Richards, 2015], where agents and humans cooperate to achieve common goals [Ramchurn et al., 2016], or human environments, where people are supported by software [Jennings et al., 2014]. In MAS, we distinguish between two groups of methods (or processes) to compose team(s), namely:

1. Exogenous Team Composition: there is an algorithm external to the agents that determines the composition of teams.
2. Endogenous Team Composition: agents themselves decide in a distributed manner the composition of a team.

Exogenous Team Composition. The team composition process uses the task requirements (i.e. constraints on teams that can be formed, such as team size [Rahwan et al., 2011a]; competences and personality as discussed in section 2.4.1) in order to build teams that are capable of solving the task with particular properties. For instance, [Crawford et al., 2016] and [Okimoto et al., 2015] consider a degree of fault-tolerance to build k-robust teams. A team is k-robust if removing any k members from the team, does not affect the completion of the task. As mentioned before, [Liemhetcharat and Veloso, 2012] propose a learning algorithm that constructs a synergy graph from observations of the performance of pairs and triples of agent. A synergy value represents how well a pair of agents work together. The authors use this learned synergy graph as well as agent capabilities to solve the team composition problem. Their method selects teams that are capable and that maximize their internal synergy.

Similarly, [Rangapuram et al., 2013] consider the competences of agents and their compatibility in order to identify a team that is both competent and compatible. Agent compatibility, expressed as a social network, can be understood as a set of preferences on team composition, such as: the inclusion of a certain team leader, or restrictions on team size, problem solving cost or agent locality (in a social or geographical sense).

In many systems, capabilities of agents are not widely known. [Chen et al., 2015] study an ad-hoc setting where agents need to co-operate with to recognize their capabilities. Agents bid for subtasks (parts of tasks) that they want to perform, though the final decision belongs to the exogenous algorithm that assigns each subtask to the best qualified agent bidding on the task.

Some approaches deal with the composition of multiple teams. For instance, [Anagnostopoulos et al., 2012] use competences and communication cost in a context where tasks sequentially arrive and teams have to be composed to perform them. Each task requires a specific set of competences and the team composition algorithm is such that the workload per agent is fair across teams. Besides the use of competences, [Farhangian et al., 2015b] use personality traits with the purpose of composing a single team.

Aside from competences and personality, team composition can also take into account agents' preferences on teams. Indeed, hedonic coalition formation employs each agent's *hedonic preferences* on its coalitions to yield a coalition structure, namely multiple teams. The defining feature of a hedonic preference is that every agent only cares about which agents are in its own team (coalition). [Spradling et al., 2013] introduce a new model of hedonic coalition formation game, the so called Roles and Teams Hedonic Games (RTHG). In this model, agents view coalitions as a number of available roles and have two levels of preferences: on the set of roles that are available in a coalition, and on their own role within each coalition.

Finally, there is relevant work on mixed teams by [Hanna and Richards, 2015], which composes a team as a pair consisting of a human and an Intelligent Virtual Agent (IVA). The pair plays a collaborative game that involves passing a sequence of obstacles to reach a target.

Endogenous Team Composition. The second group of methods for organising teams has an endogenous nature. They incorporate algorithms enabling agents to decide on team composition by themselves. In detail, agents are equipped with negotiation and decision-making mechanisms that they employ to agree among themselves on a team structure. Therefore, team composition occurs without explicit external command.

[Farhangian et al., 2015a] propose a model in which there are two types of agents: requesters in charge of tasks that seek for contributors to compose teams, and contributors that vote for the tasks they want to perform. Each requester runs an auction-based (first-price sealed-bid) algorithm with the purpose of composing teams with the highest chance to increase social wealth. Contributors issue bids pursuing to join the most useful requesters, namely the ones that are most likely to reward them. [Peleteiro et al., 2015] follow the similar approach but also employ reputation and adaptation mechanisms to allow agents in a competitive environment to autonomously join and preserve teams (as coalitions). Agents bid for tasks and each team is constructed and led by a mediator agent.

Similarly, in [Chalkiadakis and Boutilier, 2012] each agent builds its beliefs about its peers based on prior outcomes of interactions between them, and decides on coalitional actions (which coalition to join and what task to perform). Then, agents negotiate between them to form teams taking into account their own beliefs on the probability of success when being in a team.

Another interesting scenario for endogenous team composition is gaming. For instance, [Moon et al., 2005] analyse factors affecting team success in the America's Army game. There, teams can have up to fourteen players and human agents are allowed to join teams freely at the beginning of each game. However, the authors discover that the most successful team configuration has ten soldiers, moving in two sub teams (five players in each), and a long chain of communication (rather than star-shaped communication). Note that these findings actually indicate that two teams of five are more effective than one team of ten. This aligns with the team size recommendations from organisational psychology that we discuss further ahead in subsection 2.5.4.1.

There exist also mixed approches, where researchers explore both, exogenous and endogenous methods to compose teams. For instance, [Rokicki et al., 2015] propose strategies for groupsourcing (team-based crowdsourcing), ranging from team formation processes where individuals are randomly assigned to teams, to strategies requiring self-organisation where individuals participate in team building. Their results show that balanced teams (that is teams with the balanced number or agents in each team) combined with individual rewards for most effective team members outperforms the other strategies.

Analysis. The majority of researchers focuses on exogenous methods to compose teams. However, there are many actual-world application domains (e.g. co-working, or crowdsourcing) where endogenous team composition and formation are more appropriate for deployment.

Most of the literature on exogenous team composition assumes that there exists a centralized, detailed knowledge about all agents. This knowledge is required in order to compose teams based on agents' capabilities, personality, or even preferences. Endogenous methods are best for dynamic environments, where team composition and formation processes are continuously performed. Furthermore, it is a good setup for agents that learn other agents' capabilities through repeated interactions.

2.4.4.2 Team Formation

We identify two main team organisation structures to build effective teams:

1. Hierarchical; and
2. Egalitarian.

We describe each team organisation structure in the following sub-sections.

Hierarchical. A hierarchical structure considers a team leader who is responsible for and makes the decisions affecting the team. This structure is the traditional setting when it comes to business units.

As mentioned in subsection 2.4.4.1, [Farhangian et al., 2015a] consider two types of people within teams: requesters and contributors. Requesters adopt a leading function, they start a project and recruit the required people. Contributors perform the tasks assigned by requesters. The overall team behaviour is determined by the personality of agents in teams.

In [Peleteiro et al., 2015], each coalition is led by a mediator. This agent is responsible for leading a coalition by selecting suitable agents to be part of a coalition (called worker agents) and by evaluating the performance of workers while the coalition operates.

[Agmon et al., 2014] consider ad-hoc settings with two types of agents: best-response agents and ad-hoc agents. In such settings a task consists of a set of actions, and each team becomes responsible for performing a task. Each best-response agent selects its next action based on its own local world view. Each ad-hoc agent acts to bring out the best in its teammates by "leading" them to the optimal joint action. This is an arresting example of a hierarchical structure, where agents are not aware of each other's roles, and hence of a team's structure. Nonetheless, an ad-hoc agent has more knowledge than a best-response agent, and thus it exploits such information to lead its team. This may happen in a business setting, where both senior and junior staff form a team. Even though there is no clear division of roles, the senior employee uses his experience to make decisions that are best for the team in a long–term period (and may not look best from a short–time perspective).

Egalitarian. An egalitarian structure assumes that all workers in a team are equally informed and have the same rights. The leadership within a team is shared and existing team roles result from the team's task requirements. An example of this structure in real-life scenario might be a team of doctors that need to join their specialized knowledge to perform a complicated surgery on a patient.

A large part of the MAS literature focuses on the egalitarian setting, trying to benefit from leaderless teams that cooperate to complete tasks. We find this team structure in Groupsourcing [Rokicki et al., 2015], Robust Teams [Crawford et al., 2016; Okimoto et al., 2015], Ad-hoc teams [Barrett et al., 2013; Chalkiadakis and Boutilier, 2012; Chen et al., 2015], Mixed Teams [Hanna and Richards, 2015], Learning Teams [Liemhetcharat and Veloso, 2012, 2014] or Online Teams [Moon et al., 2005].

A particular case of egalitarian structure involves members that decide collectively, usually by voting, on the appropriate course of action while performing an assigned task. The real life example for this organisation structure might be a start-up with few people that make all decisions by discussion. [Marcolino et al., 2013; Nagarajan et al., 2015] and [Marcolino et al., 2016] study egalitarian structures whose agents vote to decide at every step of a task in order to choose the best course of action. They prove that teams consisting of heterogeneous agents that vote their actions are more efficient than homogeneous teams built out of the copies of the strongest agent in a team. This is because the spectrum of possible actions is wider for heterogeneous teams.

There exist also some team composition models that can produce both types of team structures. For instance, in Roles and Teams Hedonic Games model [Spradling et al., 2013], the resulting structure of the teams can be either hierarchical or egalitarian depending on the relationships between roles. Typically teams in [Rangapuram et al., 2013] are egalitarian, though the presented model includes many natural requirements that can lead to a hierarchical structure (such as inclusion of a designated team leader and/or a group of given experts).

Finally, one important question regarding team organisation requires our attention, that is, what is the effect of team network and communication structure on team performance? We already discussed the article of [Hanna and Richards, 2015], where authors show that the more informative the communication between two agents, the better the performance of the team. This result is consistent with results reported by [Sukthankar et al., 2009]. There, the authors analyse the communication patterns of teams performing a collaborative search game that simulates search and rescue scenarios. [Sukthankar et al., 2009] robustly find that the less performing teams are those that communicate less. Furthermore, [Moon et al., 2005] also highlight the importance of communication, as well as team movement structure on team performance. Regarding communication networks, two dominant communication network types are: star-shaped and long-chain shaped. Between these two, the long-chain shaped communication network performed better because it reduces team members' burden to communicate. However, the reduced communication frequency of the long-chain shaped communication network teams with respect to star-shaped communication network teams is still higher than that of losing teams. Regarding team movement, the authors found that the most effective communication network type is a dense network (team members stay close together), and that a network with two dense subgroups has fewer casualties and less communications than others but a satisfying number of opponents being killed. [Maghami and Sukthankar, 2011] introduce an agent-based simulation for exploring the effects of stereotypes on task-oriented team composition and network evolution. The authors demonstrate that stereotype value judgments can have a negative impact on task performance, even if the agents are motivated and competent enough to perform a task. Stereotype-driven agents modify the social network from which teams are formed in a systematically suboptimal way and eliminate the skill diversity required for successful task performance. [Osipov and Sukthankar, 2012] explore the relationship between network adaptation for candidate team participants and performance of problem-solving teams. Their analysis shows that the use of a higher number of skills per agent is desirable as it has a net positive effect on the number of candidate teams (where an agent can contribute its skills) and the total number of teams that can be composed by a system. However, the authors do not provide a detailed, analytical treatment of the relationship between the network adaptation policies and the teams' performance.

Analysis. The team organisation structures in the MAS literature can be grouped into hierarchical and egalitarian. The majority of MAS research focuses on egalitarian structures because of simplicity reasons. In particular, there is no need for defining a role structure together with its relationship and agent-role assignments. Although structuring teams and organisations largely helps reduce complexity of interactions, by separating responsibilities, most research in team formation does not consider a clear role division. Moreover, notice that in most business settings teams work following a hierarchical structure. Finally, research suggests that teams communicating more have higher levels of performance up to a point. However, too excessive communication leads to lower levels of performance.

2.4.5 WHEN do we do it? The dynamics

The literature on team composition and formation mostly considers that tasks are static in the sense that their requirements do not change during their execution. However, the dynamics of task arrival is considered by many. That is, there could be multiple tasks to be solved concurrently and new tasks may arrive in an asynchronous, localized manner. The different works consider different issues in this dynamic process. For instance, the number of tasks to be serviced, task and team members localization, team size per task or time limitations. Normally, if there is only one task is to be completed, the focus will be on composing the best team for the task. On a repeated task arrival setting, the use of a history of team work experiences is key to compose new teams. Hence, the literature can be classified depending on two main aspects:

1. The succession of tasks,
2. The simultaneity of tasks.

The simplest case is a one-shot task. There is neither succession nor simultaneity, and hence the problem of team composition is normally reduced to finding the best team for the only task. When tasks come in sequence without simultaneity, then the problem can be reduced to finding the best team for each task while using the learned experiences in the composition of each new team. If tasks come in succession and can be simultaneous, the need to deal with multiple teams acting at the same time becomes a key issue. The succession of possibly simultaneous tasks is the most complex framework in which memory becomes again a key element.

We discuss each aspect in detail.

2.4.5.1 Non Successive and non simultaneous tasks

In this case we face a one-shot task resolution. This is the simplest case for the team composition and formation problems. There is no long-term strategy used to compose and form teams. Thus, agents do not learn from past experiences and we cannot talk about the notion of community in this setup.

Team Composition. As mentioned above, in the team composition problem, we are looking for only one team, the best possible one to perform the task. The majority of models that consider non successive and non simultaneous tasks are simplistic. They assume that once the team is composed it has the needed attributes and will perform the task well. For instance, [Kargar et al., 2012] use agents' capabilities and team coordination cost to compose the most effective team. Similarly, [Crawford et al., 2016] and [Okimoto et al., 2015] use agents' capabilities to compose k-robust teams (see Section 2.4.4.1 for a definition of a k-robust team). In [Rangapuram et al., 2013], besides agents' capabilities, the team composition model also introduces various types of constraints (the inclusion of a specific group of agents in a team, team size, budget limitations, and maximum geographical distance between agents and between agents and tasks). This last model is more realistic, though it disregards past experiences.

Teamwork. In the teamwork phase, agents solve the task once and for all. Hence, one-shot tasks may cause self-interested behaviours, such as in [Rochlin et al., 2016]. There, as mentioned in Section 2.4.2.2, one agent (called buyer) from the team is delegated to accomplish the task of purchasing a jointly desired item with the lowest possible cost. This agent operates on a one-time setting, that is, there is a single agent deciding on behalf of the team, and hence, there is no need for that agent to behave in an altruistic manner. Authors study the notion of fairness and its influence on effectiveness. They show that the selected buyer is less motivated to do the task if the cost of the goods is to be divided equally among the team members. In this case, the purchasing costs are fully assumed by the purchasing agent. Therefore, they study different methods to reimburse the purchasing costs incurred by the buyer to improve its effectiveness.

[Hanna and Richards, 2015] study the co-operation between a human and an IVA (Intelligent Virtual Agent) in a one-shot task setting. Given that past experiences cannot be used, they experimentally show, by comparing many one-shot task instances, that the more informative the communication between the two agents, the better the performance of the team. The communication behavior of an IVA is directly related to its psychological traits.

On a different vein, many models assume that given a one-shot task, agents will behave according to their knowledge and capabilities in order to benefit the whole team. In [Barrett et al., 2013] and in [Agmon et al., 2014], team agents are pre-designed to co-operate when solving a collective task. Then, one of the agents is replaced by an ad-hoc agent that shares the team's goals, though does not know its teammates' behaviours. The ad-hoc agent cannot control its teammates, and yet it tries to improve the team's performance by learning to predict other agents' actions and thus selecting its own actions to achieve an overall optimal team behaviour. [Marcolino et al., 2013] and [Nagarajan et al., 2015] perform a one-shot task study, where team agents vote for a team action leading to the task resolution. The action voted for is sampled from a fixed probability distribution over those actions appropriate in a particular world state (no learning involved). The higher the probability of an action the more preferred it is by the agent. A plurality voting mechanism is used to select the team action. Authors show that a diverse team (with different probability distributions) can outperform a uniform team (made out of copies of the best agent) and that breaking ties in favour of the best agent's opinion in a diverse team is the optimal voting rule [2].

2.4.5.2 Non Successive and simultaneous tasks

In non successive and simultaneous tasks, the composition and formation problem becomes more complex as it now considers a set of one-shot tasks. There is still no use of the past experiences as the tasks are non successive.

Team Composition. Researchers in the area of MAS propose algorithms to compose the *best set* of teams, one per simultaneous task, instead of looking for the *best* team for a task.

In Roles and Teams Hedonic Games (RTHG) [Spradling et al., 2013] authors propose a heuristic optimization method to partition a set of agents, again to solve different instances of the same task. The method treats as votes agents' role preferences on team role structures. Firstly, the role structures of the teams will be those receiving the highest social welfare (as the summation of the agent individual utilities to play any of the roles in the structure). Secondly, the algorithm selects the agent with the highest utility for a remaining role in the most voted team role structure, recomputes the role structure preferences without that agent's preferences, and keeps staffing teams until the partition is complete. For instance, an agent may prefer to be a programmer in a two-agent team including a designer, but would not like to play any role in a team without a designer.

[2] Notice though that the authors make the strong assumption that there is a known rank of the best actions to take at any time.

Hence, an agent's role preference is not taken in isolation, but in the context of the teams' composition. Authors define Nash stable and individually stable solutions for RTHG in terms of possible local moves that agents could make within a given coalition partition and prove that every instance of RTHG has an individually stable partition that can be obtained with the use of local search movements (change of role within a coalition or coalition swaps). In our literature search, we could not find approaches dealing with different simultaneous non successive tasks.

Teamwork. Similarly to team composition, [Rokicki et al., 2015] deal with the Teamwork problem over different and simultaneous instances of the same task. Agents may change their strategy during team formation in order to reach a better solution. They classify human behaviour during team self-organisation in crowdsourcing tasks in two types. First, a number of users choose to join one of the leading teams, instead of selecting a weaker one and compete for a lower award. Second, small teams merge to form stronger teams and thus have a higher chance of achieving an award.

2.4.5.3 Successive and non simultaneous tasks

When tasks are successive and non simultaneous, the algorithms for team composition and formation deal with a task that has to be assigned to a team, and in many cases solved, before new tasks arrive. A successive setting can discover phenomena which we believe are important, but which are not captured when the attention is limited to static, non successive tasks. If in the system of the same set of agents, teams are created and dismantled depending on the task, the agents may behave very differently than in a non successive settings. For instance, a person will behave in a different manner if she repeatedly borrows a car from her friends, than when she simply rents a car. The successive setting has its advantages: it lets agents learn from the past experiences and build their beliefs based on this knowledge.

Team Composition. In [Anagnostopoulos et al., 2012], the first task arrives at the first time step and is assigned to a newly composed team of experts before the arrival of the second task. This procedure repeats until all tasks are assigned. Authors propose an algorithm to compose a set of teams to handle a set of these incoming tasks. The goal is to form a new competent team upon arrival of each task, so that the workload in the whole system is balanced. There is no learning involved in this process. Contrarily, in [Liemhetcharat and Veloso, 2012] a learning algorithm is proposed that constructs a synergy graph from observations of the performance of pairs and triples of agent in solving previous tasks. The synergy tells how well a pair of agents work together and they use this learned synergy graph as well as agents' capabilities to solve the team composition problem for the next task. Their method selects teams that are capable and maximize their internal synergy.

Teamwork. To the best of our knowledge, there are no contributions on teamwork that consider successive and non simultaneous tasks.

2.4.5.4 Successive and simultaneous tasks

When tasks are successive and simultaneous, the algorithms for team composition and formation deal with a set of tasks arriving, possibly overlapping in time that have to be assigned to newly composed teams.

Team Composition. In [Farhangian et al., 2015a], tasks arrive in any order, possibly overlapping in time. A team is composed for each incoming task and after execution agents assign performance values to each one of the other team members. These values are public and used by the community to compose teams for future tasks. [Chalkiadakis and Boutilier, 2012] present several learning algorithms to approximate the optimal Bayesian solution to the repeated team composition. Similarly, [Peleteiro et al., 2015] compute, after teamwork, both individual agent and coalition (team) reputation values to be used in the composition of future teams.

Finally, in [Chen et al., 2015], for each new task arriving agents decide which team to join balancing exploitation (rewards from completing tasks learned from previous task solving) and exploration (learning opportunities from more qualified agents leading to future rewards).

Teamwork. To our knowledge, there are no contributions considering successive and simultaneous teamwork.

Analysis. One time settings (i.e. non successive tasks) are usually simplified models that do not take into consideration the history of agent interactions. One-shot tasks may cause self-interested behaviours, where agents look for at least a fair split of costs associated with teamwork. However, the majority of the literature on team composition and teamwork considering this setting assume that the agents will always behave accordingly to their capabilities and knowledge. The successive tasks provide us with more realistic and complex scenarios. The tasks arrive either in order, one after another, or overlapping in time. The majority of the literature uses this setting to let agents build their beliefs based on the past experiences and compose new teams according to these beliefs. Regarding teamwork, there are no contributions that explore successive settings. In other words, the state of the art does not acknowledge the memory of agents as important while executing tasks.

2.4.6 WHERE do we do it? The context

The context is understood as the circumstances that form the setting for the team composition and formation processes. We observe that the concept of context in the reviewed computer science literature has not played a major role so far. Contrarily, according to the organisational psychology literature [Guzzo and Dickson, 1996], it is one of the most important variables while composing and forming teams (see Section 2.5.6). There are different categorizations of context. One of them is proposed by [Kozlowski and Bell, 2013], which classifies contexts as follows:

- Organisational Context: technology used, organisation structure, leadership, culture, and climate.
- Team Context: normative expectations, shared perceptions, and compatible knowledge (generated by and emerge from individual interactions).
- Individual Context: attributes, interactions, and responses.

In the MAS literature there are very few works that consider the social context while composing teams. [Terveen and McDonald, 2005] set a framework for social matching systems, which aims to bring people together on both physical and online spaces. They explain the importance of context in recommending a member of social network for collaboration. In [Rangapuram et al., 2013], while composing teams, the context is exemplified as a social network that encodes the previous collaborations among experts. The idea behind it is that the teams that have worked together previously are expected to have less communication overhead and work more effectively as a team. Similarly, [Peleteiro et al., 2015] propose to express social context by the reputation measure. There, upon task completion, the contractor rates the quality of the service provided by a team and, also teams rate their own workers. Finally, this rating information is maintained and aggregated by a reputation module. [Liemhetcharat and Veloso, 2012] propose to model a social context by using the learned synergy graph (that measures how well agents work with one another) and hence, solve the team composition problem. [Anagnostopoulos et al., 2012] include the coordination costs by means of a social network over the set of agents and assume a metric distance function on the edges of the network. On top of modeling preferences based on social context (such as past interactions, compatibility in collaborating, distance in a company's hierarchy), the function may include any other kind of context, (for instance geographical proximity between agents or between task and agents within a team).

Analysis To the best of our knowledge, there are only few works in MAS literature that recognize the context as an important variable. Besides [Anagnostopoulos et al., 2012], which considers both social and geographical contexts, the methods in the literature only consider the social context (if analyzed at all).

2.5 Team engineering in Organisational Psychology

In this section we discuss all above aspects in detail answering the questions asked in the introduction of this chapter.

2.5.1 WHO is concerned?

First, we are going to survey the literature on Organisational Psychology that deals with the attributes of humans composing teams. We discuss further methods to measure human attributes in Chapter 3.

We will use the structure as in section 2.4.1.

2.5.1.1 Capacity.

In OP, the most important capacity of team members that is related to team performance is their cognitive ability. Hence, the main goal is to study how cognitive abilities influence team performance. Cognitive ability refers to the 'capacity to understand complex ideas, learn from experience, reason, solve problems, and adapt' [Devine and Philips, 2001, p.507]. Hence, cognitive ability in OP is a much wider concept than capacity in multiagent systems as on top of skills widely used in MAS systems, it contains many other attributes such as experience, competences, age or even gender.

Moreover, in contrast to computer science, where capabilities are static, psychologists deal with the dynamism of human capacity. Humans learn new capabilities and increase their level every day for whole live (see more in [Laal and Salamati, 2012, p.399-403] for the concept of the lifelong learning). There are diverse tests and methods to examine humans capacity, such as: intelligence or cognitive competences tests, assessment centers or social and behavioural competence tests.

Regarding team composition, on the one hand [Bell, 2007] and [Devine and Philips, 2001] found that mean team values of cognitive ability are correlated with team performance. Moreover, she also found that the lowest and the highest team members' cognitive abilities are correlated with team performance in lab and field settings. In addition, [Devine and Philips, 2001] found that the variance of team members' cognitive ability did not help predict team performance. These authors also found that the mean value is twice more informative in predicting than the lowest and the highest member's scores. On the other hand, [Devine and Philips, 2001] found that cognitive ability influences team performance differently depending on contextual variables (such as working normative procedures or human resources policies). These findings suggest that, when composing a team, organisations and managers should not only take into account the members' cognitive ability, but also the context in which the team will operate. This will be further discussed in Section 2.5.6.

[Woolley et al., 2015] discuss the existence of a measurable collective intelligence in teams that is analogous to individual intelligence. Authors suggest the existence of a general collective intelligence factor that explains a team performance on a wide variety of tasks. [Woolley et al., 2010] show that collective intelligence is correlated with the average social sensitivity of group members, the equality in distribution of conversational turn-taking, and the proportion of females in the group. In STEM (Science, Technology, Engineering and Math) teams, gender diversity can enhance group processes, which are increasingly important as collaboration becomes a centre piece in the production of science. The enhancement of group processes and higher levels of collective intelligence can, in turn, lead to greater innovation and scientific discovery [Bear and Woolley, 2011]. Finally, similarly to findings in the computer science literature, the concept of team properties is normally understood as a sum of humans' individual attributes.

2.5.1.2 Personality

In addition to the before-mentioned individual attributes, the literature has examined the role of personality. The most prominent approaches have been the "Big Five" personality traits theory [Mount et al., 1998], Schutz's theory of fundamental interpersonal relations orientation (FIRO) [Schutz, 1958] and the Myers Briggs Type Indicator method [White, 1984]. They have been used to find the personality traits and types associated with team performance. Regarding the "Big Five" theory, meta-analytic research has found that certain levels of conscientiousness, openness to experience and agreeableness are good performance predictors [Mount et al., 1998].

Another approach is that of the theory of fundamental interpersonal relations orientations (FIRO) [Schutz, 1958]. The idea is that humans have several needs (i.e. need for inclusion, control and affection) and that groups with team members that have compatible needs will perform better than those with incompatible ones. Nevertheless, research has found mixed support for this theory [West, 2012a].

Some companies have also tried to base their team formation on cognitive styles of the members, by using the Myers-Briggs Type Indicator (MBTI) assessment instrument —[Myers et al., 1998], which is a questionnaire that measures cognitive styles along four dimensions: Extraversion — Introversion, Sensing — Intuition, Thinking — Feeling, and Judging — Perceiving. Nevertheless, there is not enough rigorous research evidence showing its relationship with team performance [West, 2012a].

There are also novel approaches created with the purpose of team composition and formation. For instance, the Post-Jungian Personality Theory, which is a modified version of (MBTI) [Wilde, 2013]. It operates on the same dimensions as MBTI. The main novelty of this approach is its use of the numerical data generated by the instrument [Wilde, 2011]. The results of this method seem promising as within a decade this novel approach tripled the fraction of Stanford teams awarded national prizes by the Lincoln Foundation [Wilde, 2009]. However, the method is not yet properly validated and tested, which makes it disregarded by psychologists.

2.5.1.3 Analysis.

Several correlations have been found between cognitive ability and team performance. The personality is also present while composing teams, although the correlation between personality and team performance is not clearly explained.

The most widely used test to measure personality is the "Big Five". Organisational Psychology studies show that besides cognitive ability and personality, experience and gender are further attributes to consider for team composition [West, 2012b]. Indeed, research findings on this topic suggest that diversity in those characteristics can have an effect on team performance and innovation [West, 2012b]. Additionally, some further research has also paid attention to values and has found collectivism and teamwork preferences [3] to be additional good team performance predictors [Bell, 2007].

2.5.2 WHAT is the problem?

When it comes to team composition, the organisational psychology literature has focused on defining task classifications. These classifications have been employed to study the relation between task types and team performance. Hence, in this section we will review the most known task classifications and its influence on team performance.

Two of the most widely discussed task classifications are those of [McGrath, 1984], [Hackman and Lawler, 1971; Hackman, 1990] and [Hackman and Oldham, 1975]. While the classification of [McGrath, 1984] is based on the cognitive requirements of tasks, the classification in [Hackman and Lawler, 1971; Hackman and Oldham, 1975; Hackman, 1990] is based on the motivation characteristics of tasks (i.e. autonomy, task variety, task significance, task identity and task feedback). The research on team composition show that the classification based on the motivation characteristics predicts more accurately the team performance [Podsakoff et al., 1997].

[Hackman, 1990] defines a task classification based on motivational requirements composed by seven work task types:

1. top management;
2. task force;
3. professional support task;
4. performing task;
5. human service task;
6. customer service task;
7. production task.

The classification of [McGrath, 1984] based on cognitive requirement proposes three dimensions that characterize each task type:

1. Choose-Execute;

[3] Teamwork preferences refer to team members preferences on other team members to work with.

2. Conceptual-Behavioral;
3. Conflict-Cooperation.

Technically speaking each task type becomes a 3-tuple with qualitative values for each dimension. For instance, a routine task would be very executive, medium behavioral and low conflicting.

After analyzing seventeen classifications in the literature [Wildman et al., 2012] came out with a different classification as follows:

1. Managing others;
2. Advising others;
3. Human service;
4. Negotiation;
5. Psychomotor action;
6. Defined problem solving;
7. Ill-defined problem solving.

As an alternative perspective, [Navarro et al., 2011] propose a task classification based on the task context (namely task complexity, interdependencies between subtasks in a task, and uncertainty about the dynamics of the environment where the task is executed and the lack of information). Their results show that in order to achieve acceptable performance, the greater the complexity, interdependence and uncertainty, the stronger the requirements on the maturity of teams (e.g. joint experience, cohesion) and on the diversity of team members' capabilities. For instance, to carry out highly interdependent tasks, all team members should possess coordination skills (maturity) and some of them the capacity to take decisions (diversity). Taking into account other task context characteristic (i.e. uncertainty and interdependence) their study results show, the greater the uncertainty and interdependence of task types, the more diverse the competences for team members to cope with complexity. From the other hand, if the team is overqualified for the task to perform, the motivation of team members decreases and the quality of the outcome is lower or the task is not completed at all.

2.5.2.1 Analysis.

The OP literature provides many different classifications of task types, where the most important are the classifications based on the motivation of individuals, the cognitive abilities and the task context. Provided the amount of classifications and the apparent lack of consensus among them, we believe that choosing among the several classifications previously presented in order to apply them to the study of team composition is a hard decision. Nevertheless, such decision must be made in order to move forward with the understanding of how a task type can influence team composition. In an attempt to advice researchers, notice that the research show that the classification based on the motivation characteristics predicts more accurately team performance.

From OP perspective team performance cannot be assessed by simply measuring how long it takes for the group to finish a certain task or by counting the number of right answers to predefined and clear questions, which is a common approach in computer science. OP rather analyzes joint team objectives and the team composition and formation setting (such as not realistic deadlines, a number of individuals in a team, the level of stress in a team or the quality of the outcome).

The current research on organisational psychology focus has moved from task analysis so not many results are present. Although task types are defined, different task instances constantly appear because of technological development. That makes it very difficult to keep the pace. That is why the focus on OP moved to competences (understood as cognitive ability, see Section 2.5.1). This is why not much work has appeared after defining task taxonomy. At the same time task complexity increased and hence, teams are getting more and more important. Moreover, a clear mapping between cognitive ability of individuals and task types is needed. As a major benefit such mapping would ease team composition.

2.5.3 WHY do we do it?

In OP the main objective for team composition and formation is to maximize team performance. When measuring it, the research on OP suggests that we should go beyond mere economic criteria, the quality of decision-making processes or other traditional performance indicators [Hackman, 2002; Komaki, 1997].

An important difference with respect to the computer science literature is that team performance is considered from two perspectives: objective and subjective. On the one hand, objective team performance refers to the features of the outcome of a team (e.g. quality, delivery time, cost, sustainability). On the other hand, subjective team performance refers to the quality of human resources in a team (e.g. motivation, satisfaction, commitment, illness rate, stress) [Quijano et al., 2008]. Therefore, while the first one refers to the delivered output of a team (what customers obtain), the latest one focuses on the inner development of team members. Objective and subjective team performance are significantly correlated (e.g. [Quijano et al., 2008]). Therefore, and not surprisingly, the organisational psychology literature considers both types of performances when tackling team composition and team formation (e.g. [Meneses and Navarro, 2015]). The subjective and objective performance of a team are determined by the several aspects of the context (discussed in Section 2.5.6), together with individual characteristics, the task and the team processes. Following [Navarro et al., 2011] the subjective and objective performance of a team are determined by the adjustment between the maturity level of the team (e.g. in terms of group development, potential, etc.) and the groups tasks characteristics.

Analysis. An important difference with respect to the computer science literature is that team performance is considered from two perspectives: objective and subjective. Objective and subjective team performance are significantly and directly correlated. Therefore, and not surprisingly, the organisational psychology literature considers both types of performances when tackling team composition and team formation. The computer science literature can benefit from the concept of subjective team performance that currently disregarded. Therefore, current team composition models, which mainly focus on the objective team performance, need to be extended.

2.5.4 HOW do we do it? The organisation

Similarly to Section 2.4.4 on computer science, we divide the organisation into two aspects: team composition and team formation.

2.5.4.1 Team Composition.

The organisational psychology research on team composition has been very influenced by task classification. For several authors, there is a relationship between task type and team type (structure). For example, according to [Hackman, 1990], there are seven team types based on the task type to perform:

1. top management;
2. task force;
3. professional support;
4. performing groups;
5. human service;
6. customer service;
7. production teams.

[Devine, 2002] and [Delgado Piña et al., 2008] highlighted that team performance depends on a good matching between team type and task type.

On the other hand, there are multiple team type classifications in the literature based on other criteria [Devine, 2002; Gibson and Kirkman, 1999; Marks et al., 2001]: motivation-based, cognitive-based or context-based (see section 2.5.2), though none of them has been widely used or accepted. Also, there is agreement that team diversity must be exploited while composing teams. Diversity refers to the degree or level to which the members of a group differ or contrast in one or more attributes. Diversity has been shown to have an impact on team performance [Mathieu et al., 2008]. In their review, [Mathieu et al., 2008] point out the vastness of the literature featuring team diversity and draw attention to four main diversity dimensions: demographic, personality, functional background, and attitudes and values.

[Horwitz and Horwitz, 2007] conducted a meta-analysis to understand the relationships between team diversity and team performance. For this, they differentiated between two classes of diversity: bio-demographic and task-related. The former refers to diversity in individual attributes that are observable and not learned (e.g. personality, gender, age, ethnicity), whereas the latter regards diversity in acquired capabilities, such as education or expertise. Using meta-analytic techniques, they found task-related diversity to be positively correlated to both qualitative and quantitative measures of team performance. However, they did not find a clear relationship between bio-demographic diversity and team performance. Although pointing out the small number of studies supporting these latest findings, their preliminary results seem to give more importance to the diversity of acquired team member attributes, such as the type of education or knowledge expertise.

Finally, another factor influencing team performance is team The relationship between team size and productivity is a question of broad relevance across economics, psychology, and management science. Hence, the size of a team is one of the most frequently studied parameter when analyzing team performance. There is a disparity in the literature due to the fact that appropriate team size is dependent on the task and the social context in which the team operates. When it comes to athletics, sport teams have a defined number of team players: A football team needs 11, the Standard Platform League in RoboCup five players per team, and baseball teams require nine players. But when it comes to organisations, it is hard to find a golden rule to determine the optimal number of team members. For complex tasks, however, where both the potential profits and risks of teamwork increase with the number of team members, neither theoretical studies nor empirical evaluations consistently favor larger vs. smaller teams [Mao et al., 2016]. Regarding established theories, psychology [Steiner, 2007], economics [Holmstrom, 1982], and management [Malone and Crowston, 1994] studies suggest that increasing team size can be harmful to team performance. This happens because: individuals find it tempting to free ride on the efforts of teammates [Holmstrom, 1982; Steiner, 2007]; the overhead associated with communication increases with team size [Steiner, 2007]; and communication among team members causes partitioning into sub-teams [Lorenz et al., 2011] and chitchat [Tetlock et al., 1992]. Therefore, in complex tasks, where all these reasons may exist simultaneously, the relationship between team size and performance is not well described by existing theories. [Mao et al., 2016] performed a study of the dynamics of team performance and its relationship with team size in the digital volunteer setting of crisis mapping. Their findings show that although social loafing and coordination costs result in reduced contribution from individuals in larger teams, the potential benefits of coordination can outweigh this loss in performance.

However, other studies show that there is an inverse relationship between the size of the team and its performance [Bartol, 1977; Oyster, 1999]. [Oyster, 1999] and [Bartol, 1977] show that team size is important when analyzing team performance. Yet, they have offered different recommendations concerning the best size for various types of tasks to achieve acceptable performance. [Oyster, 1999] states that the right number of people in a team depends on the kind of tasks team members need to perform. They believe that for teams ranging from four to six, all the team members' competences can be fully used, but for larger teams some members' competences are under-used and this provokes that teams split up. According to the studies of [Bartol, 1977], the optimal number of members for problem-solving tasks is five. He states that there is a limit to the team size, which, if exceeded, causes a drop in the performance of the team. [Bartol, 1977] says that in the case of a team containing more than six people there is a tendency to split the team into two, which brings about negative effects. The cause is twofold: high coordination cost and loss of motivation by team members [Oyster, 1999].

Finally, some studies have found team size to be unrelated to performance [Martz et al., 1992] or that increasing team size actually improves performance without limit [Campion et al., 1993].

2.5.4.2 Team Formation.

Once a team has been composed, there are different processes that the team carries out to execute the task and achieve the collective goal. Several classifications of team processes have been proposed in the literature, from which, the most recent and overarching one is the one proposed by [Marks et al., 2001] and [Goodwin et al., 2009; Salas et al., 2005]. Typically the research investigated the ways of implementing team processes and of measuring how well teams perform. To begin with, [Marks et al., 2001] distinguish between three broad types of processes: action-orientated, transition-orientated and interpersonal. The first ones refer to actions that team members undertake to accomplish goals, namely team monitoring, systems monitoring, monitoring progress towards goals and coordinating activities. Regarding transition-orientated processes, these are actions related to planning and/or evaluating in order to guide in attaining team goals, that is goal specification, mission analysis, formulation and planning, and strategy formulation. Finally, interpersonal processes are those intended to manage interpersonal relationships. They comprise motivating/confidence building, conflict management and affect management [Marks et al., 2001]. On the other hand, [Salas et al., 2005] built upon previous research and narrowed down the main processes into "Big Five" team processes: team orientation, backup behaviour, team leadership, adaptability and mutual performance monitoring.

Another important aspect is that team climate influences the effectiveness of processes. A team climate is defined by the degree to which a team of people possesses certain core attributes that are needed for the team to work effectively. These attributes include the interrelationship among team members, the identification of each person with the team and its social values, the coordination of team resources, behaviours and technologies, as well as the desire of each team member to achieve the objectives of the team [Meneses and Navarro, 2015]. A good climate assures the sharing of resources, mutual rewards and information exchange. It promotes a high level of openness, safety, and a mix of upward, downward and horizontal communication processes that help to increase team performance [Knapp, 2010; Kozlowski and Ilgen, 2006; Mathieu et al., 2007; Rico et al., 2010].

A team climate that is conductive to learning requires shared perceptions of work settings [Brodbeck, 2003; James et al., 2008; Ramirez-Heller et al., 2014]. According to [Brodbeck, 2003] and [Ramirez-Heller et al., 2014], a team climate conductive to learning is characterized as one in which:

1. There is empathy, support, as well as a common understanding among its members, conveying an atmosphere of mutual trust,

2. There is a regular contact as well as informal and formal communication processes among its members,

3. There exists a common agreement with the goals and objectives to be achieved, and these shared goals are clear, realistic and feasible,

4. There is a prevailing notion of democracy and equality among its members, with no one having particular control over the others,

5. Members perceive a personal development as the team enhances their creativity and provides general support in fulfilling their individual plans.

Finally, there are various studies in Organisational Psychology analyzing the effect of communication and network structure on team performance. Typically, teams in organisations are strategically composed by heterogeneous individuals [Osatuyi, 2012]. This is based on the assumption that once team members share their information, the team as a whole can access a larger pool of information, knowledge and expertise. However, studies have shown that teams, unlike individuals, sometimes do not effectively share and use the unique information available to them. This leads to poorer decision making. Informational influence theory holds that the subjective importance of information may affect if information is shared or not. Henceforth, an important factor for performance improvement is the proactive communication of information about team members' goals [Butchibabu et al., 2016]. It is also found that task complexity is negatively correlated with information exchange. Surprisingly, teams tend to share less information when working on complex tasks, compared to when working on simple tasks [Osatuyi, 2012].

Also in an on-line game domain communication plays an important role on the performance of virtual team members [Leavitt et al., 2016]. For instance, League of Legends enables non-verbal communication through "pings," alerts that are easy to activate and provide auditory and visual hints for teammates. [Leavitt et al., 2016] analyse 10.293 matches in this popular game and test the impact of ping actions on team performance. They show that pings by players have a positive but concave relationship with player performance. That is, teams sending more pings have higher levels of performance up to a point after which sending more pings leads to lower levels of performance.

Another important factor influencing team performance is team shared belief in their collective power to produce desired results [Yildir, 2005]. In [Yildir, 2005], the team shared beliefs of computer game players were measured as 126 teams competed in a highly interdependent, online role-playing team game. Structural equation modeling results indicated that for all interdependent teams, as team shared belief increased, both team persistence and performance also increased positively and linearly.

Analysis. Regarding team composition, there is a strong relationship between task type and team type (structure). The type of the team depends on the features of the task to perform and so very often team types are derived from task types. Besides task type, team diversity plays an important role when composing teams. Regarding the "optimal" team size, it is a complex question and future research is needed to determine the impact of team size on team performance, such as the nature of the task, the internal motivations, and the context. Some preliminary results show that the more complex the task, the larger the size of the team needs to be, but limited to an optimal size of six members. Regarding team formation, several different team processes classifications have been proposed, though no agreement has been reached. Finally, having a good team climate seems key to achieve good performance.

2.5.5 WHEN do we do it? The dynamics

Humans learn with every interaction. Our memory recollection and capability improvement cannot be removed or stopped. Hence, the organisational psychology research usually deals with complex scenarios, those of simultaneous and successive tasks, see Section 2.5.5. In organisational psychology, the dynamic attributes of a team are referred to as emergent states. Emergent states develop during teamwork and have an effect on the outcomes. Several examples of emergent states [Mathieu et al., 2008] are team confidence, team empowerment, cohesion, team climate, collective cognition or trust between team members.

The development of emergent states is closely connected to the process of team learning behaviours. As members of a team interact with one another and perform tasks, they learn from their experiences. That is, they learn by asking questions, seeking feedback, experimenting, reflecting on results, and discussing errors or unexpected outcomes of previous actions [Edmondson, 1999]. These complex tasks allow team members to acquire, share, combine and apply knowledge [Kozlowski and Ilgen, 2006; Olivera and Argote, 1999]. They also lead to the development of shared understanding and meaning as well as to the acquisition of mutual knowledge, skills, and performance capabilities [Garavan and McCarthy, 2008]. All these developments enhance team performance [Edmondson, 1999; Zellmer-Bruhn, 2006].

Analysis. Unlike computer science, the reviewed organisational psychology literature does not study simple scenarios such as non successive and non simultaneous tasks. Typically, organisational psychology analyzes complex and realistic scenarios as human learning capabilities need to be considered. Moreover, on top of including the social network and memory about the outcomes of past experiences, the researchers in organisational psychology deal with the dynamics of individuals' capabilities (as humans learn new capabilities and forget not used ones).

2.5.6 WHERE do we do it? The context

From a systemic perspective teams are part of the structure of an organisation and therefore they operate within this organisation. In the same way, an organisation is part of the environment. The environment creates demands and requirements for an organisation and influences the organisation's system. In turn, the organisation tries to address these requirements by influencing the operations of its teams and their performance in diverse ways.

Research results suggest that context plays an important role in the performance of teams [Guzzo and Dickson, 1996; Hackman, 1990]. [Hackman, 1990] between others propose and analyse many contextual factors that have to be considered when composing a team:

- The uncertainty on the level of complexity of the tasks and the degree of dynamics of the environment. Both aspects influence the uncertainty within the organisation and therefore its teams need to operate with incomplete knowledge. The uncertainty about external factors is determined by the available information about the customers, the suppliers, or other competing organisations. The uncertainty about internal factors is determined by the dynamics of tasks, organisational rules and objectives. In such an uncertain context, teamwork is more challenging and paradoxically teams may perform better than in a stable and predictable context.

- The vision and mission of an organisation that determine the main rules and norms to be followed and what is to be considered as good performance.

- The set of values, policies and strategies of the organisation. For instance, organisations supporting individual values will hinder teamwork and team performance will thus be poor. This is because teamwork is based on shared values, mutual support, constructive collaboration, mutual trust, coordination mechanisms and synergies, which are collective values. On top of it, an organisation promoting internal competition will lead to individual strategies of withholding information and self-interested behaviours.

- The organisational benefits such as the reward or the training systems. Diverse motivational theories are available to explain the relevance of the reward systems for increased performance. For example, teams will perform better with an appropriate reward system.

- The resources and assistance made available to the team. It is obviously easier for the team to achieve good performance when operating in a context of resource abundance.
- The organisational climate. A context with a perceived climate of control and low level of autonomy for the team will hinder successful teamwork and performance. As teamwork requires an individual engagement with the team, a climate is needed that facilitates information sharing or team skills development.
- The cultural context. The definition of a team changes across cultures: in cultures valuing individualism teams are seen more as a set of people each contributing to a different subtask, whereas in cultures valuing collectivism teams are seen as having shared goals, values and responsibility for the whole task. Research results show that teams perform better in a collective cultural context.

Analysis. In contrast with computer science approaches, the context where teams solve tasks plays an important role in the organisational psychology literature. The context is understood as internal and external factors influencing teamwork. The internal context can be characterised as dimensions of the organisation, such as vision and mission, values, policies and strategies, or organisational benefit system. The external context can be characterized as dimensions of the environment in which the organisation operates, that is the culture, the available resources, and the uncertainty about other players behaviour.

2.6 Discussion

Computer Science (CS) and Organisational Psychology (OP) have followed rather disparate approaches when it comes to team composition and team formation. However, some similarities and differences can be drawn and several new research questions can be formulated from a cross reading of the two literature corpus. In Table 2.3 a comparison of the main papers in CS can be found.

2.6.1 Similarities in both approaches

When modeling agents' attributes in CS, there are two main approaches. There is either extensive a-priori information about teammates given as input or ad-hoc scenarios where agents learn their teammates' capabilities. In OP a number of tests are proposed to acquire a-priori information about teammates, such as intelligence or cognitive competences tests, assessment centres or social and behavioural competence tests. Also, similar to CS, OP studies allow to learn human capabilities from their repeated interactions.

To maximize team performance, one of the crucial findings in both OP and CS is that team members have to be heterogeneous.

Regarding the tasks that are executed by agent teams, both OP and CS focus rather on team members' attributes required to perform a task than on a detailed planning of the task execution.

2.6.2 Differences in both approaches

The first difference we find between CS and OP is with respect to the complexity of individual team members. Organisational psychology focuses on humans with all their intrinsic complexity while CS focuses on a limited set of human-like attributes to build software agents. In CS the agent attributes have been categorized as personality and capacity. In OP, although human attributes can also be categorized as personality and capacity, capacity is a much wider concept. It contains not only skills, but also other attributes, such as competences, experience, gender or age. Moreover, while in OP the human capabilities are assumed to be dynamic (i.e. lifelong learning), software agents capabilities are assumed to be static and only the behaviour model may change with agents' interactions.

In CS the majority of approaches assume that the joint capabilities of agents in a team are enough to solve a given task. However, the researchers in OP recognize also other factors as important when composing and forming a team, such as the motivation of individuals and the task context. They also show that the motivation characteristics predict more accurately the performance of a team than the other factors. Regarding OP research gaps, it lacks a mapping between cognitive ability of individuals and task types (which is an input in CS models) which complicates team composition.

Ref.	**Team Process**	**Individual Properties**	**The task**	**The Objective**	**Team Composition**	**Team Organisation**	**The dynamics**	**The context**
Agmon et al., 2014	Formation	Personality	Plan-based	Maximizing social welfare	Exogenous	Hierarchy	Non Successive and Non Simultaneous	N/A
Anagnostopoulos et al., 2012	Composition	Capacity	Individual-based	Maximizing the quality	Exogenous	Egalitarian	Successive and Non Simultaneous	Social, Geographical
Barrett et al., 2013	Formation	Personality	Plan-based	Maximizing the quality	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A
Chalkiadakis and Boutilier, 2012	Composition	Capacity	Individual-based	Maximizing social welfare	Endogenous	Egalitarian	Successive and Simultaneous	N/A
Chen et al., 2015	Composition	Capacity	Individual-based	Maximizing social welfare	Exogenous	Egalitarian	Successive and Simultaneous	N/A
Crawford et al., 2016	Composition	Capacity	Individual-based	Minimizing cost	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A
Farhangian et al., 2015a	Composition	Personality	Individual-based	Maximizing social welfare	Endogenous	Hierarchy	Successive and Simultaneous	N/A
Farhangian et al., 2015b	Composition	Capacity and Personality	Individual-based	Minimizing cost	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A
Hanna and Richards, 2015	Formation	Personality	Individual-based	Maximizing the quality	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A
Kargar et al., 2012	Composition	Capacity	Individual-based	Minimizing cost	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A

Table 2.2: Comparison of the computer science contributions reviewed in this chapter.

Ref.	**Team Process**	**Individual Properties**	**The task**	**The Objective**	**Team Composition**	**Team Organisation**	**The dynamics**	**The context**
Liemhetcharat and Veloso, 2012	Composition	Capacity	Individual-based	Maximizing the quality	Exogenous	Egalitarian	Successive and Non Simultaneous	Social
Marcolino et al., 2013	Formation	Personality	Plan-based	Maximizing the quality	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A
Nagarajan et al., 2015	Formation	Personality	Plan-based	Maximizing the quality	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A
Marcolino et al., 2016	Formation	Personality	Plan-based	Maximizing the quality	Exogenous	Egalitarian	Successive and Non Simultaneous	N/A
Okimoto et al., 2015	Composition	Capacity	Individual-based	Minimizing cost	Exogenous	Egalitarian	Non Successive and Non Simultaneous	N/A
Peleteiro et al., 2015	Composition	Capacity	Individual-based	Maximizing the quality	Endogenous	Hierarchy	Successive and Simultaneous	Social
Rangapuram et al., 2013	Composition	Capacity	Individual-based	Maximizing the quality	Exogenous	Egalitarian / Hierarchy	Non Successive and Non Simultaneous	Social
Rochlin et al., 2016	Formation	N/A	Plan-based	Maximizing the quality	N/A	Hierarchy	Non Successive and Non Simultaneous	N/A
Rokicki et al., 2015	Composition, Formation	N/A	N/A	Maximizing social welfare	Exogenous, Endogenous	Egalitarian	Non Successive and Simultaneous	N/A
Spradling et al., 2013	Composition, Formation	Capacity	Individual-based	N/A	Exogenous	Egalitarian	Non Successive and Simultaneous	N/A

Table 2.3: Comparison of the computer science contributions reviewed in this chapter.

The CS literature has focused on team co-operation with various objectives that can be categorized as at least one of the following: minimizing overall cost, maximizing social utility, or maximizing the quality of the outcome (understood as maximizing team performance). In OP, the main objective for team composition and formation is just to maximize team performance. Moreover, from an OP perspective team performance cannot be assessed by the time spent to perform a task, by comparing costs or by counting the number of right answers as it would ignore some important subjective reasons. Instead, OP analyzes possible causes of failure, such as an excessive amount of work needed to execute the task given the size of the team or the lack of motivation of team members. This is why the performance is assessed from two perspectives: objective and subjective, while, CS only considers objective measures. In CS there are only early attempts to include a subjective perspective while analyzing team performance. It is shown that the motivation increases by introducing competition mechanisms (like in crowdsourcing teams) or by giving agents freedom while selecting their collaborators (like in ad-hoc teams).

Since in CS agents can be modeled depending on the needs, researchers can study different settings depending on the dynamics of task arrival (one task or many, one time or many). Many MAS models are simplistic since they consider only one task arriving at a time. Unlike CS, the reviewed OP literature does not study simple scenarios, since humans have memory and improve their capabilities with every task. Hence, typically OP analyzes only complex and realistic scenarios. The CS literature uses these complex scenarios to let agents build their beliefs based on past experiences and compose new teams according to these learned beliefs. OP, on top of including the social network and memory about the outcomes of past experiences, deals with the dynamism of individuals' capabilities (as humans learn new capabilities and forget not used ones).

2.6.3 Cross fertilization opportunities

Prior sections explored a range of concepts and issues concerning team composition and formation. In this final subsection, we focus on posing research questions for the field, organised around a set of research opportunities:

- **Establish a connection with the OP literature.** We pose the following research questions:

    1. **What criteria to use when composing effective teams?** A goal of OP is to improve organisational performance by placing the right people in the right jobs, thus enhancing the fit between the individual and the organisation. This includes manual methods for building effective teams. Nevertheless, research on team composition and team formation in CS and OP has evolved separately. The MAS literature has typically disregarded significant OP findings, with the exception of several recent, preliminary attempts (like [Farhangian et al., 2015a] or [Hanna and Richards, 2015]). This body of research has focused on algorithms that help automate team formation and

composition. Research findings from the OP literature have much potential for MAS heuristics (such as team diversity [Mathieu et al., 2008], team size [Mao et al., 2016] or context [Guzzo and Dickson, 1996]).

2. **Are current CS methods enough to measure team performance?** From an OP perspective, team performance cannot be assessed by simply measuring how long it takes for a group to finish a certain task or by counting the number of right answers to predefined and clear questions, which is a common approach in CS. OP rather analyzes joint team objectives and the team composition and formation setting (such as unrealistic deadlines, the number of individuals in a team, the level of stress in a team or the quality of the outcome). Also, OP focuses on the inner development of team members and analyses the quality of human resources in a team, that is, motivation, satisfaction, commitment, illness or stress rate [Quijano et al., 2008]. When evaluating team performance, Computer Science research should take into account team objectives, task dependencies, the feasibility of the task, etc.

3. **How to exploit the factors that influence team performance?** According to OP research, in order to carry out highly interdependent tasks, all team members should possess coordination skills (maturity) and some of them the capacity to take decisions (diversity). Also, the greater the uncertainty and interdependence of task types, the more diverse the competences for team members to cope with complexity. However, if the team is overqualified for the task to perform, the motivation of team members decreases and the quality of the outcome is lower or the task is not completed at all. All these dependencies have been studied extensively by OP research, but they are ignored by CS. We should work to understand what is the correlation between task type and team type and what is the exact influence on team performance.

- **Enhancing agent models.** The CS literature is in need of analysing more complex examples where agents are modeled as humans. Based on our findings we form several research questions for MAS research:

  1. **How to develop richer information (or cognitive) agent models to enhance team composition?** In OP, the most important capacity of team members that is related to team performance is their cognitive ability. It is a much wider concept than the notion of capacity in multiagent systems, since beyond skills, widely used by MAS research, it contains many other attributes such as experience, competences, age, or even gender. While some of the human attributes may not make sense in an agent context (like age or gender), some do (such as cognitive abilities, lifelong learning or behavioral model). Also, there is a need to include more sophisticated models for agent

capabilities, such as graded capabilities instead of binary ones. Richer agent models would allow the CS field to further benefit from OP findings for team composition and formation.

2. **How to model and exploit competence dynamics?** The majority of CS models assume that competences are a fixed attribute of each agent. OP indicates that human capabilities are necessarily dynamic (evolve along time) so that teams can successfully perform tasks in dynamic real-world scenarios and in a variety of contexts. The dynamics of competences through learning and experience and the cultural values could be used by MAS research to program adaptive agents, specially when interacting in mixed teams involving humans.
3. **How can we include agents' motivation in team composition and formation models?** OP research highlights motivation as an important factor for team performance [Hackman, 1990]. The majority of the MAS literature on team composition and teamwork assumes that agents always behave according to their capabilities and knowledge. While in MAS research it is shown that motivation increases by introducing competition mechanisms (like in crowdsourcing teams, [Rokicki et al., 2015]), or by giving agents freedom when selecting their collaborators (like in ad-hoc teams, [Agmon et al., 2014]), these are only early attempts to include agents' motivation as an important factor for team performance.

- **Enhancing task execution.** We are interested in the following research questions for multiagent research:

  1. **Are agents' joint capabilities enough for successful task execution?** Regarding the tasks that are executed by agent teams, CS focuses on those team members' attributes required to perform a task rather than on a detailed planning of task execution. The majority of approaches assume that the joint capabilities of agents in a team are enough to solve a given task. There are some preliminary attempts to include planning, though they are very simplistic. The majority of methods do not consider time constraints, action dependencies, action failure, plan robustness, task dynamic changes and hence, the vast literature on planning has not yet been integrated into team formation methods.
  2. **How to endow agents with competence learning capabilities?** Since in CS agents can be engineered depending on the needs (i.e. agents can be designed with different attributes, such as personality or memory, depending on the whole system design), researchers can study different settings depending on the dynamics of task arrival. The CS literature uses complex scenarios to let agents build their beliefs based on past experiences and compose new teams according to these learned beliefs. However, while executing tasks, there are no contributions that explore successive or simultaneous settings. Agent

learning when executing tasks could be used to further improve the task execution.

- **Enhancing team performance through context inclusion.** Particularly, we are interested in the following question:

    1. **How to computationally exploit the context within team formation and composition?** OP research results suggest that context plays an important role in the performance of teams, [Guzzo and Dickson, 1996; Hackman, 1990; Terveen and McDonald, 2005]. Although, to the best of our knowledge, there are only a few works in CS that would recognize context as an important factor, besides the social and geographical context considered in some papers. There is a need to perform further research on how to computationally model the context within team composition and team formation to build better performing agent teams.

- **Enhancing team modeling** We form the following research question for multiagent research, that is:

    1. **Is the sum of the agents' individual capabilities enough to predict team performance?** Although individuals' attributes have been extensively studied and considered, there is still a need for modeling the global properties of agent teams. Such modeling should go beyond considering simple properties such as the sum of the agents' individual capabilities or the Boolean representation of whether the team can perform a task or not. One of the findings from OP that could be used is a general collective intelligence factor that explains team performance on a wide variety of tasks, [Woolley et al., 2010].

Chapter 3

Individual Profiling Model

In Chapter 2 we analysed the literature from Organisational Psychology and identified individuals' attributes that influence team performance. In this Chapter we describe the dominant approaches in Organisational Psychology, Industrial Psychology research, and Human Resources practices and summarise their major findings when it comes to tools to measure attributes of individuals that can be useful in a team composition processes. In other words, in this Chapter, we discuss methods for *Initial Assessment* process as a part of our management organisational workflow presented in Chapter 1 (shown in figure 1.1).

In theory, the general idea is pretty straightforward. When one knows what makes the members of a team effective, and in which combination these attributes work best, it is possible to use this knowledge to compose high-performing teams. At the basis of such selection must be sound empirical evidence that the team member attributes in question are related to measurable team performance. However, collecting such evidence is easier said than done. Nonetheless, considerable work in fields such as organisational psychology, and industrial psychology has focused on various factors that influence team performance [Arnold and Randall, 2010; Mount et al., 1998; West, 2012b; White, 1984]. These factors include competences, experiences, age and gender as well as personality. While some of these are straightforward to collect by a self-evaluation form, such as age or gender, others like personality or competences are more difficult to measure.

The remaining of this chapter is organised as follows. In Section 3.1 we discuss main approaches to measuring individual personality. Next, in Section 3.2 we describe main approaches to evaluating individual competences. Finally, in Section 6.5 we summarise our main findings.

3.1 Personality

Personality determines people's behaviour, cognition and emotion. Different personality theorists present their own definitions of personality and different ways to measure it based on their theoretical positions. The most explored schemes to measure personality are using subjective self-assessment questionnaires that are called *personality tests*. Typically, the outcome of a personality test consists of several personality dimensions that define the individual. In what follows, we refer to these outcomes of the personality tests as *personality traits*. We divide personality theories to compose teams into two approaches: Individual Traits Approach and Team Balance Approach. In this section, we discuss these approaches in detail.

3.1.1 Individual Traits Approach

The most explored approach is based on the presumption that, when it comes to predicting a team's performance, some individual personality traits matter more than others. An issue with this approach is that personality traits of team members are individual-level concepts and team performance is a group-level concept. Henceforth, researchers who take the individual approach must develop group-level concepts of individual traits and then must investigate their relationship with team performance. Henceforth, researchers in Organisational Psychology are interested in the following questions [Chiocchio et al., 2015]:

1. Does a team perform better when it has a high overall level of the trait in question?

2. Does a team perform better when team members are diverse on a particular trait?

The majority of researchers distinguish between two characteristics of team composition in terms of personality, that is *the elevation* and *the variability* of a certain trait within a team [Peeters et al., 2006]. Trait elevation is an average or a sum of individual values for a trait, or the proportion of individuals with a high value on a trait (ibid). Suppose, for instance, that it was found experimentally that teams with the higher elevation of trait X perform better than teams wherein the elevation of trait X is lower. The indication for team member selection here is clear: select team members so that the trait elevation is the highest.

Trait variability is represented by a team's variance or standard deviation for a certain trait (ibid). In other words, the relevance of a particular trait for team performance may lie in the way in which it varies across the team members and not in its overall level in the team. Suppose, for example, that it is shown experimentally that teams whose variability of trait Y is high perform better than those teams in which the variability of trait Y is low. This, too, has indications for team members selection. In this case, when selecting team members it is important to include the ones that differ greatly when it comes to trait Y.

One of the tests that received attention in the literature is the Myers-Briggs Type Indicator (MBTI) scheme designed to indicate psychological preferences in how people perceive the world and make decisions [Briggs and Myers, 1980]. It consists of four dichotomous dimensions on a binary scale, that is:

- Extraversion vs Introversion (E–I),
- Sensing vs Intuition (S–N),
- Thinking vs Feeling (T–F), and
- Judging vs Perceiving (J–P).

Within this approach, every person falls into one of the sixteen possible combinations of the four letter codes, one letter representing one dimension (see figure 3.1 for details).

Figure 3.1: Sixteen MBTI personalities. The figure comes from OEC^2 Solutions (2018)

Each type is said to specify a set of behavioural tendencies, reflecting differences in attitudes, orientation, and decision-making styles [Boyle, 1995]. This approach is easy to interpret by non-psychologists, though reliance on dichotomous preference scores rather than continuous scores excessively restricts the level of statistical analysis [Devito, 1985]. These psychometric limitations raise concerns about the validity of the instrument. Additionally, test-retest estimates raise doubts about the stability of MBTI-type scores [Boyle, 1995].

Currently, the dominant model in Organisational Psychology literature appears to be the Five Factor Model (aka FFM [Costa and McCrae, 1992] or "Big Five" [Goldberg, 1990]), which uses five broad dimensions to describe human personality. This model was developed using two different methodologies, but converged on the same five factors of personality [Peeters et al., 2006]. That is:

- Extraversion — refers to the degree to which an individual is socially active, open and talkative,
- Agreeableness — refers to the extent to which an individual is polite, trusting and cooperative,
- Conscientiousness — refers to the degree to which an individual is achievement-driven, diligent, and organised,
- Emotional stability — refers to the extent to which an individual is low on anxiety and anger,
- Openness to Experience — openness to experience refers to the degree to which an individual is curious and imaginative (see also figure 3.2 for further description of FFM traits).

According to psychologists, every individual can be described in terms of these five traits and they remain relatively stable over time and across situations (ibid.).

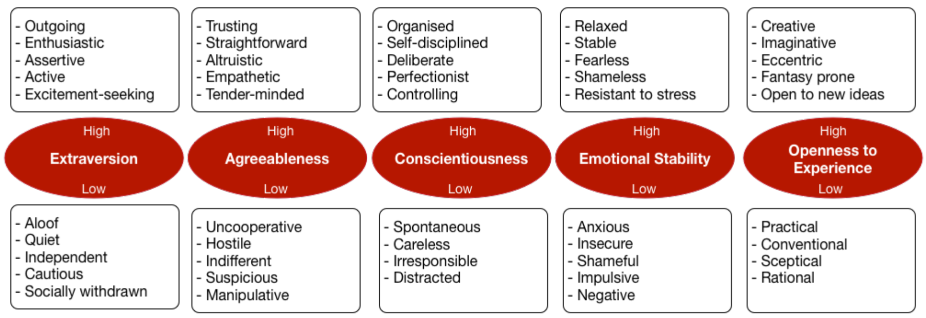


Figure 3.2: The personal attitudes consisting of five factors in the Five Factor Model.

Two recent meta-analytic studies support the importance of some of personality traits in team composition represented as an average of values for all team members (here, referred as team means). Bell [Bell, 2007] found that for each of the Big Five personality traits examined separately, team means were positively correlated to team performance. Prewett et al. [Prewett et al., 2009] examined all but Openness to Experience trait and reported similar patterns.

Both Bell and Prewett et al. also investigated the relationship between team member variability and team performance with respect to each of the traits separately. In both studies, variability effects on performance were generally weaker than mean effects. Bell showed some evidence that variability with respect to Conscientiousness as well as Openness to Experience could be problematic for field teams. Prewett et al. suggested that variability with respect to both Conscientiousness and Agreeableness could be problematic, but only for reciprocal tasks (in which work is circulated back and forth among team members) [Tesluk et al., 1997]. However, sample sizes for all these findings were relatively small and it is unclear whether these results will hold in future research [Peeters et al., 2006].

Mohammed and Angell [Mohammed and Angell, 2003] present contradictory results. They examined student project teams whose task was to improve processes based on issues identified in organisational settings. The researchers measured Agreeableness, Conscientiousness, Extraversion, Emotional Stability, and Team Orientation of all teams using the team mean and team variability. Interestingly, none of those traits, when considered separately, was meaningfully connected to team performance.

Additionally, despite the popularity of the Big Five in recent years, its construct validity has been questionned [Boyle, 2008; Jang et al., 2002; Toomela, 2003]. Toomela [Toomela, 2003, p. 723] reported that a coherent FFM personality structure emerged only among samples of individuals who had received extensive formal education, thereby raising doubts as to the genetic determination of the postulated Big Five personality dimensions. While two factors (Extraversion and Opennes to Experience) appear to be universally accepted and they appear in all major contemporary models of broad personality traits [Zuckerman et al., 1993], the other three Big Five dimensions (Openness to Experience, Agreeableness, and Conscientiousness) continue to remain controversial [Boyle, 2008]. Moreover, according to Poropat (2002), Big Five personality instruments fail to detect significant gender differences in personality structures. It is also argued that the Big Five dimensions are too broad and heterogeneous, and lack the specificity to make accurate predictions in many real-life settings [Boyle, 2008]. Finally, to our knowledge, there are no contributions in organisational psychology literature that have a clear team composition method based on this scheme. Organisational Psychologists give some directions regarding the elevation of some traits, although it is not clear what levels make a "good team" and if these traits are indeed needed by all team members.

Finally, taking an Individual approach to team composition seems counterintuitive. Perhaps one of the main reasons why personality seems relevant is a need for compatibility among team members. In other words, team members should "fit together" in order for the team to achieve its potential. Some personality types may work well together, while others might not. Therefore, in the next section we discuss Team Balance approaches taken by organisational psychologists to compose teams.

3.1.2 Team Balance Approach

In Team Balance Approach researchers try to understand which team member attributes are best in terms of the configuration that they compose. Here, the question is not whether the team's mean on a single trait influences team performance. Instead, this approach explores how team members fit together with respect to traits of individual members of the team.

Surprisingly, research examining team balance approach to personality is quite limited. One of the first theories that gained popularity was the Fundamental Interpersonal Relations Orientations (FIRO) theory proposed by Schutz [Schutz, 1955]. It is based on the idea that there are three human needs relevant to teamwork that need to be balanced within a team, that is:

- Need for inclusion,
- Need for control, and
- Need for affection.

According to Schultz's theory, teams whose members have balanced needs are the most effective. That balance is imposed by matching individuals that have high levels of different needs with individuals that have low levels of those needs. The reason for this matching given by Schultz is that an individual who is high on one need can only have that need satisfied by an individual who is low on the same need. For instance, if everyone on a team was high on "need for control," then there might be many internal conflicts as all team members took positions of leadership. As a consequence, team performance would likely suffer. In 1958, Schutz developed the FIRO-B survey to assess proposed needs, which theoretically could be used for team composition purposes. However, the support for this theory is rather weak [Hill, 1975; Moos and Speisman, 1962; Shaw and Webb, 1982]. Moreover, Hill reported that teams whose members were considered incompatible, using Schutz's approach, actually performed better than those judged to be compatible. Interestingly, the literature review done by Chiocchio et al. (2015) did not uncover recent empirical studies on the FIRO-B and team member compatibility, suggesting that this theory has been discarded by psychologists.

Another theory proposed by Belbin emphasises the importance of *roles* in team composition processes [Aritzeta et al., 2007]. In essence, Belbin claims that there are nine required team roles that need to be balanced for an effective team. These roles include: plant, resource investigator, coordinator, shaper, monitor evaluator, implementer, teamworker, specialist and completer–finisher. The description of roles is shown in figure 3.3.

According to this theory, most people have a number of "preferred team roles" that they naturally display. They also have "manageable roles" that are roles which might not be the most natural course of behaviour for them, but they can display them if required by the situation. Finally, people have least preferred roles, those they should not try to perform. In this last case, the effort is likely to be great, and the outcome, poor. However, there is no fixed number of roles for each person. Because of humans displaying multiple roles, a team of three or four may potentially cover all nine roles [Bell, 2007]. A team is considered balanced (and, thus, theoretically, a high-performance team) when at

Figure 3.3: Nine Belbin roles needed for an effective team.

least one member has a high score on each role [Senior, 1997]. Belbin proposes a questionnaire to measure team roles called "Belbin Team Role Self-Perception Inventory" [Belbin, 1993].

Similar to the FIRO-B issues raised earlier, there is a lack of strong evidence supporting Belbin's theoretical arguments [Chiocchio et al., 2015]. One major problem is that the test used to assess roles does not have convincing psychometric properties [Anderson and Sleap, 2004; Broucek and Randell, 1996]. Although some limited support for the theory has been reported in studies with very small samples (e.g. 10 teams in Senior, 1997), the Belbin roles tend not to be related to team performance [Batenburg et al., 2013; Partington and Harris, 1999; van de Water et al., 2008].

Another theory in Team Balance approach is the Post-Jungian Personality Theory [Wilde, 2009, 2013]. Its author, Wilde, suggests that both a diverse personality profile of team members and a balanced gender distribution, positively influence the effectiveness of a team. Here, effectiveness is understood as the probability of goal achievement while performing problem-solving tasks. He proposes a novel method that is a modified version of the Myers-Briggs Type Indicator (MBTI) [Briggs and Myers, 1980], the "Step II" version of Quenk, Hammer and Majors [Wilde, 2013]. The questionnaire to determine personality is short, contains only 20 quick questions (compared to the 93 MBTI questions [Boyle, 1995]). This is very convenient for both experts designing teams and individuals doing the test since completing the test takes just a few minutes (for details of the questionnaire, see [Wilde, 2013, p.21]). In contrast to the MBTI measure, which consists of four binary dimensions, the Post-Jungian Personality Theory uses the *numerical* data collected using the questionnaire [Wilde, 2011]. Douglass J. Wilde claims that it covers the same psychological territory as MBTI [Wilde, 2009]. He also suggests that the numerical data obtained through an MBTI questionnaire can be used as an input for team composition.

Similarly to MBTI, the test is based on the pioneering psychiatrist Carl Gustav Jung's cognitive mode personality model. It has two sets of variable pairs called psychological functions:

- **Sensing / Intuition (SN)**: The sensing function S "includes all perceptions by means of the sense organs" [Jung, 1971], whereas the intuition function N "is perception by means of the unconscious" (ibid).

- **Thinking / Feeling (TF)**: By the thinking function T Jung means "...intellectual cognition and the forming of logical conclusions," whereas "feeling (F) is a function of subjective valuation" (ibid).

and psychological attitudes:

- **Perception / Judgment (PJ)**: The attitude energy for information collection (P) is independent of and usually different from that for decision making (J) [Jung, 1971].
- **Extroversion / Introversion (EI)**: Extroversion is the flow of psychic energy outward toward the exterior world: "an outward turning of libido" [Jung, 1971], whereas introversion draws psychic energy towards one's interior psyche: "an inward turning of libido" [Jung, 1971].

Psychological functions and psychological attitudes compose together a personality. Every dimension of personality (EI, SN, TF, PJ) is tested by five questions. Each question can be answered in three different ways. The full questionnaire is presented in Appendix B. As an example, table 3.1 presents questions to measure the Extrovert / Introvert dimension.

EI1	You are more:	(e) sociable	(i) reserved
EI2	You are more:	(e) expressive	(i) contained
EI3	You prefer:	(e) groups	(i) individuals
EI4	You learn better by:	(e) listening	(i) reading
EI5	You are more:	(e) talkative	(i) quiet

Table 3.1: The part of the questionnaire to collect the EI dimension

Let us take for instance *EI1*. A user can select "sociable", "reserved" or both answers. The numerical value of each dimension is calculated as follows. Take again the EI questionnaire, we calculate the number of (e) answers (those at the left in the table), subtract from them the number of (i) answers and normalise by dividing by 5 which is the number of questions. The result is then a value in $[-1, 1]$. We repeat this procedure for each dimension to get a vector of four values (EI, SN, TF, PJ) $\in [-1, 1]^4$.

To construct teams, Wilde gives a set of guidelines that are slightly different for each team size. However, in general, the rules can be summed up as follows:

- SN and TF personality dimensions should be as diverse as possible within a team;
- A team should have at least one student scoring positive on EI, TF and PJ dimensions, namely an extrovert, thinking and judging person (called ETJ personality);
- A team should have at least one student scoring negative EI dimensions, namely an introvert; and
- A team should be balanced in gender.

In summary, in this subsection we discussed the Team Balance Approach that tries to find compatibility between different personalities and based on that, the best team members configurations (on the contrary to the Individual Approach that claims that some individuals are better than others). Curiously, the research examining this approach is limited and the most known tests are widely criticised. Hence, in the next Chapter we explore the Post-Jungian Personality Theory, a novel method proposed by Douglas J. Wilde to compose effective teams. This method was never thoroughly tested, however the provided results of this method are promising, since within a decade this novel approach has tripled the fraction of Stanford engineering student teams awarded national prizes by the Lincoln Foundation [Wilde, 2009].

3.2 Competence

A single competence is defined as a set of behaviors representing one aspect of what is needed to perform a given task [Washington and Griffiths, 2015]. For instance, *Decision Making* is frequently cited as a personal competence associated with a cluster of behaviors that include assessing the importance, risk and urgency related to each situation and taking actions which are in the best interest of the organisation [Parker and Fischhoff, 2005]. A complete collection of competences defined in an organisation is called a *competence model*. Competence models provide means for human resources for individual processes (such as recruitment, promotion, evaluation, and training) as well as for group processes (such as team building, co-learning, composing proficient teams or estimating the probability of project success).

However, many times, different (and sometimes conflicting) competence models are used for distinct human resources processes within the same organisation. Recruiters use one set of competences, project managers employ another, and couches adapt a third set as learning objectives. On top of that, once the human resource process is completed, the collected information about competences is rarely reused by another process. This leads to inefficient and many times unfair human management systems that rather confuse employees than help them improve.

The solution for this problem of conflicting competence models is to adopt a *common set of competences* that can be *reliably measured* to support all human resources processes across the organisation. Imagine having one reliable competence model that serves many different purposes at the same time. First, it helps to define a set of competence requirements for positions that we recruit for. Second, it provides a set of criteria to compare candidates' competences with position needs. Third, it helps clarify individual role requirements, thus setting expectations for performance management. Fourth, it provides competences as learning objectives in development programs. Finally, it helps in performance appraisals by serving as a rubric and thus, keeping them more objective. This way, a common model is introduced, which helps in reliable assessment of employees competences, and thus, makes employees understand organisational values, clears expectations from them, and proposes consistent learning and development plans.

Henceforth, in this section we focus on two aspects of competences. First, we discuss approaches to developing a competence model in an organisation. Second, we describe methods to measure competences once the model is created centering around two different human resources processes, i.e. recruitment and progress tracking.

3.2.1 A Competence Model

Having a consistent competence model in an organisation is extremely useful in a broad range of applications, such as recruitment efficacy, training and development, workforce assessment, etc. It provides the benchmark by which all the employees know what is expected, and how well they are meeting the organisation's needs. There are three main approaches that can be taken to develop an organisational competence model:

1. **The Job Competence Assessment Method (JCAM).** JCAM [Dubois and Rothwell, 2004] uses interviews and observation of both outstanding and average performers working on the same positions to determine which competences distinguish between those two groups. According to the authors this method leads to the development of a valid and reliable competence model when the method is carefully applied. It consists of the collection and analysis of data obtained through *behavior event interviewing*. There, the interviewers ask employees to describe both successful and unsuccessful work experiences. Typically, researchers require between 6 and 12 individual interviews for each position. Based on the collected data three sets of competences are identified [Dubois and Rothwell, 2004]:

    - those of the exemplary performers (that distinguish performance),

  - those of both exemplary and average performers (the minimum worker competences),
  - those of the average but not the exemplary performers (discarded as the exemplary performance do not show them).

  Although JCAM produces rich and comprehensive work-related data, it has limitations. First, it is very related to analysed jobs. The collection of competences for all positions could serve as a global competence model, but it would require a deep analysis of all positions within an organisation. Second, it requires competent interviewers and statistical support services, which is very costly. Third, key employees must be available for interviews, which is a time-consuming process.

2. **Developing a Curriculum approach (DACUM).** In contrast to JCAM that takes into account only position holders, DACUM relies on all work experts (i.e. performers, supervisors, and possibly customers if they are highly informed about the position) [Dubois and Rothwell, 2004]. The experts describe the tasks performed daily and this information becomes the basis for discovering the competences essential for the positions. DACUM requires less work than the JCAM model, although it has similar limitations, i.e. it is still a very time-consuming and costly process.

3. **The generic approach.** The most popular approach is to use one of the models previously developed by industrial and organisational psychologists or organisational practitioners. This is the cheapest method to build an organisational competence model. However, competence models from external sources may be of questionable quality as their source is typically not known. Therefore, we should carefully examine the origins of the competence models we want to use. Additionally, the generic competence models must be tailored to meet the needs of an organisation (depending on the values and objectives of a company as well as the positions' specifications). While there are too many competence models available to be detailed in this work, we refer the reader to Washington and Griffiths [Washington and Griffiths, 2015] citation for an example of the model and additional information on this topic.

3.2.2 Competence Assessments

Once we have a consistent competence model defined, we need to establish ways to measure competences of each employee. In this subsection we provide some ideas from the literature on how to measure employees' competences so that the process is transparent, fair and efficient.

3.2.2.1 Cognitive Assessments

In psychology research, the most common method to measure abilities of people is to perform a cognitive assessment, i.e. either a set of puzzles of the cognitive capabilities or a self-assessment questionnaire measuring a variety of aptitudes. They are an inexpensive and efficient means of data collection as it is possible to test large numbers of people in a relatively short period of time. Literally thousands of tests measuring different competences are now available that can be used by organisations as shown by a simple Google search. These can be used when reassured that they are valid and reliable, i.e. they measure the competences that they are supposed to measure and generate results consistent among employees (for instance a measuring tape is a valid tool to assess a person's height and it is reliable when comparing people's heights). However, implementing a new competence model and testing current employees against all competences that were defined in an organisation is still a challenging, difficult and time-consuming process. To facilitate it and quickly obtain an initial idea of employees' competences, organisations can test their employees against general mental abilities.

The most traditional approach is measuring an intelligence quotient (IQ) [Noruzi and Rahimi, 2010]. However, whilst IQ tests measure a variety of different abilities such as mathematical, spatial, linguistic and reasoning, people tend to have certain abilities that standard IQ tests fail to recognise [Noruzi and Rahimi, 2010]. For instance, the capacity to maintain a good relationship with stakeholders is very important in organisations, yet it is not covered by a "general intelligence" model.

Howard Gardner, in his Multiple Intelligences theory, claims that human intelligence has multiple dimensions. He defines intelligence as "the capacity to solve problems or to fashion products that are valued in one or more cultural settings" [Gardner and Hatch, 1989]. For over two decades Gardner studied brains of individuals who suffered a brain injury (such as impairment or paralysis) [Noruzi and Rahimi, 2010]. He compared their brains with healthy people brains and he found that the disabled brains are damaged in specific areas. This way he discovered seven parts of the brain responsible for the specific physical functions and he associated them with seven (later on, eight) different intelligences (See figure 3.4).

The intelligences relate to an individual's unique set of competences and ways they demonstrate intellectual abilities. According to the author all intelligences are equally important and consist of [Carter, 2009]:

1. Verbal-linguistic intelligence — The capacity to manipulate language effectively and to express oneself, whether in writing or orally. This intelligence involves the sensitivity to the phonology or sounds of language, the meanings or semantics, words and the practical uses of language. It also accounts for the use of language as means to remember information.

2. Logical-mathematical intelligence — ability to think logically, conceptually and abstractly, reason deductively and detect logical and numerical patterns. It also involves sensitivity to logical statements, relationships and propositions (if-then, cause-effect), and other related abstractions.

3. Spatial-visual intelligence — capacity to perceive world in images and pic-

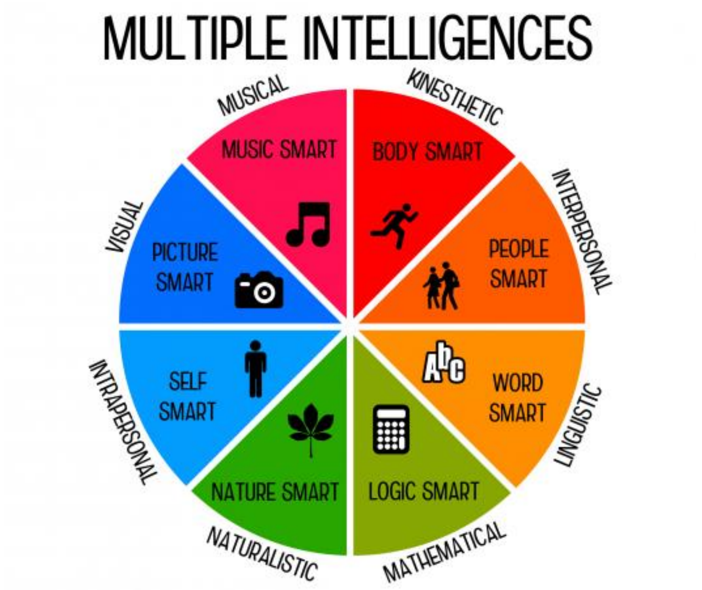


Figure 3.4: Multiple Intelligences of Howard Gardner. The figure comes from [Kunesh, 2018].

tures and to accurately and abstractly represent visual or spatial ideas. It includes sensitivity to shape, color, line, space, form, and the relationships between these elements.

4. Bodily-kinesthetic intelligence — capacity to control one's body movements, to use body to express ideas and feelings and to handle objects skillfully. It involves physical abilities such as coordination, dexterity, balance, flexibility, strength, and speed.

5. Musical intelligence — ability to produce and appreciate rhythm, pitch and melody. The ability to perceive, transform, and express musical forms.

6. Interpersonal intelligence — capacity to perceive and respond appropriately to the moods, intentions, motivations, desires and feelings of others. It includes sensitivity to facial expressions, gestures and voice as well as the ability to influence a group of people to follow one's desires.

7. Intrapersonal — ability to have an accurate picture of oneself (one's strengths and limitations) and in tune with inner moods, intentions, feelings, values, beliefs, motivations and desires. It includes the capacity for self-discipline, self-esteem, and self-understanding.

8. Naturalist intelligence — ability to recognize and categorize numerous species of fauna and flora. This also involves sensitivity to other natural phenomena (e.g., mountains, cloud formations, etc.). This intelligence

was a later addition to the model and is not as widely accepted as the other seven.

There are various questionnaires developed to measure this model. For instance, Rice [Rice, 2013] proposes suitable tests for different human age intervals (i.e. $6-7$, $8-14$ and 15+) that vary in complexity (the higher the age interval, the higher the complexity). For this reason, we have performed an experimental study in secondary education using this theory, where we used the test suitable for teenagers of ages between 12 and 14 (please see Chapter 5 for details of the experiment and Appendix A for complete Intelligences test).

Next, we divide human resources processes into two subcategories, i.e. Recruitment and Performance Tracking. We categorise competence assessments' methods based on these processes and we discuss each method.

3.2.2.2 Recruitment Process.

The most important decision organisations make regarding people management is who to employ as a bad hire can be very costly [Holmes, 2013]. Competence models can help in this crucial decision by establishing position criteria. That is, using the competence model recruiters can select a subset of competences desired for a given position. Based on that, they can define job requirements, create an ideal candidate profile, test candidates for the selected competences and compare candidates against those competences. For instance, if troubleshooting is part of a support analyst position then testing the competence of problem solving and decision making in a candidate's history would help verify that competence. Assessing candidate fitness in terms of competences is most frequently accomplished by:

- **Work Samples.** Some organisations ask candidates during the recruitment process to fill in competence quizzes or perform work sample tests that are similar to tasks that she can face during the real job. These tests are based on the premise that the best predictor of future behavior is observed behavior under similar situations [Deb, 2006]. Work samples can serve as a great source for an initial assessment of employees competences.

- **Employee Achievement History.** The recruiters typically ask candidates to provide them with proofs of previous activities such as the results of academic and development programs, professional certificates, portfolios, peer-reviewed articles, white papers, knowledge exchange tools and platforms, video presentations, demos, technical instructions, working projects, git repositories, webpages as well as recommendation letters. All these sources can be used as part of an assessment to measure the competence level of the potential employee. However, it is important to understand the exact contribution of an employee in the presented work as well as the the information verity as candidates have a tendency to glow over their past.

- **Competence-based Interviews** Interviews are the most common procedures to assess candidates' competences. Practically, all selection processes use one or more interviews [Anderson et al., 2001]. While telephone screenings are more efficient in terms of time, face-to-face interviews are more appropriate when in-depth information is needed (such as competence assessment) (ibid.). Typically, an interview consists of a set of verbal open-ended questions [Gusdorf, 2008]. Hence, the interview is really a verbal test for a candidate. However, unlike a paper and pencil test, the results are subject to interpretation by the interviewer(s) and thus can have a huge potential for error. Nonetheless, research suggests that building an interview on an organisational competence model greatly increases interview efficacy [McDaniel et al., 1994]. This technique requires the recruiter to be competent enough in the competence it's interviewing for. Without this knowledge, evaluation may vary greatly from one recruiter to another. To improve the efficacy even further some organisations use either a set of one-to-one interviews or a panel of interviewers [Ryan et al., 1999]. A set of one-to-one interviews is especially used by big tech companies such as Amazon, Google, Facebook, Microsoft, etc. There, each interview is a new opportunity for a candidate to show her competences. The aggregation algorithms of interviewers' opinions vary from one company to another and they are not openly communicated to the public. When it comes to a panel of interviewers, it typically consists of the hiring manager, human resources representative, and the experts assessing specific competences. The panel asks interview questions and all panel members hear the responses and independently judge the candidates. The data coming from competence-based interviews is a time efficient, inexpensive and relatively reliable source for employees' competence assessment. Additionally, in an ideal scenario all employees should pass a similar recruitment process, therefore the collected data should be unified and complete (besides some extraordinary cases like mergers or acquisitions).

- **Assessment Centres.** Assessment centres have been a best recruitment practice since the 1950s [Washington and Griffiths, 2015]. The Standards and Ethics for Assessment Center Operations [Rupp et al., 2015] define an assessment centre as a process that "consists of a standardized evaluation of behavior based on numerous inputs. Multiple trained observers and techniques are used. Judgments about behavior are made, in part, from specifically developed assessment simulations". According to Washington and Griffiths the set of activities varies from one assessment centre to another and it may include:

  - Leaderless group discussions (a group of candidates is instructed to engage in a discussion on a given topic and no candidate is designated as a leader),
  - In-basket simulations (candidates receive a number of mails, documents and phone calls and they have a limited period of time to set

priorities and organise their working schedule accordingly),
- Business case analyses and presentations,
- Role-plays.

The methodology behind the assessment centres is very rigorous. For instance, some research claims that there have to be at least three trained assessors for each participant [Washington and Griffiths, 2015]. Thanks to that, assessment centres can make a better selection decision and are great predictors of employees' competences. Unfortunately, not many companies use assessment centres because of their high cost and even if they do, to the best of our knowledge, once the process is finished the data is not reused.

3.2.2.3 Performance Tracking Process.

The success of a prosperous and sustainable organisation comes from the ongoing development of a competent workforce [Washington and Griffiths, 2015]. There are various methods that can be introduced in organisations to track employees' performance and development. Those include:

- **Training Activity.** Once an organisation has a valid competence model, individual competences that are common to a given role or level (e.g., entry-level supervisors) easily translate into learning objectives [Washington and Griffiths, 2015]. For example, many entry-level supervisor training programs teach key competences such as Problem Solving, Active Listening, Time Management, and Communication (ibid.). These programs use a variety of instructional techniques including lectures, exercises, case studies, and situation simulations. There are also workshops focused on a single competence, for instance a Public Presentation. In these typical few-days workshops, participants may attend a lecture on the elements of the ideal public speech, videos of good and bad speeches to compare and contrast and typically they are asked to deliver speeches, give and receive feedback. The employees' progress can be observed during the workshop or it can even be tested at the end of the event to assess newly acquired competences.

- **Peer-assessments.** The initial competence set can be gathered (or objectified if we already have the competence information) using peer assessment appraisals. In the conventional performance appraisal or review process, a manager periodically (normally annually) writes her assessment of the performance of a reporting employee (ibid.). This is the simplest and least costly solution, although allowing significant biases. These biases can go both ways — employees can benefit or be punished by the manager personal likes and dislikes [Buckingham, 2011]. Firstly, managers tend to remember the most recent events instead of analyzing the entire year's performance. Secondly, the importance of an initial impression might heavily influence an appraisal irrespectively of subsequent performance. Finally, personal bias can come from a manager's views about race, nationality, gender, religion, age, disability, hair color, intelligence, etc. Therefore, some companies try to lower the importance of biases by collecting assessments from various sources.

One of the most widely used tools is the 180/360 multi-rater feedback method [Barth and de Beer, 2017], where feedback of peers, self-reviews or even direct reports of clients are included to help evaluate an employee's true competences. These surveys have become a best-practice assessment tool, used in the majority of large American organisations [Washington and Griffiths, 2015]. As mentioned before, asessments are subjective by nature, although having multiple sources makes the aggregation less subjective. According to Washington and Griffiths, the current 360 evaluation process used in organisations goes as follows:

- Employees identify and invite their peers to give an anonymous feedback of employees' competences. The peers are limited to those who have cooperated with the employee long enough to be able to assess their competences. Typically, at least a dozen of peers is selected and online surveys are sent to them.
- The surveys request the assessments to be real numbers on a predetermined scale to allow the normative comparisons of perceived competence levels of employees across different peer groups (direct managers, team colleagues, direct reports, the employee herself, etc.). More robust surveys also request open-ended written feedback to justify given assessment numbers. The assessments are aggregated with respect to a type of reviewer (direct managers, peers, direct reports, the employee herself, etc.), using simple or weighted mean of all given assessments (like in systems such as Hudson (uk.hudson.com), Success Factors (successfactors.com), Halogen Software (halogensoftware.com), Appraisal-smart (appraisal-smart.com), WLH Consulting (wlhconsulting.com) and many more).
- *A feedback report* is created from the survey results and a qualified coach delivers and interprets the report for employees to assure that they understand it.

Multi-rater appraisal is a good method for an initial measurement of employees' competences. However, it is not ideal for a number of reasons. Firstly, it focuses on rating a person's performance in a given period of time. These appraisals are too broad and too subjective, making the collected data biased [Buckingham, 2011]. Secondly, the number of reviewers required for this assessment method needs to be relatively high. Therefore, some organisations find it too expensive to collect and process that amount of data every year or half a year. Nevertheless, the 180/360 process could be less costly if it was smartly introduced in an organisation. For instance, code reviews or integration tests can be a good opportunity to evaluate programming, architecture or design skills of engineers. Finally, the information of potential biases could be included in the final assessment by allowing peers to comment on the assessments of others. In Chapter 6, we present a new ranking algorithm that can be used to evaluate employees' competences. It uses peer opinions as well as peer judgments over those opinions (i.e. a second level evaluation) to detect biased reviewers and diminish the importance of their opinions by the usage of a reputation measure.

3.3 Summary

In this chapter we described the dominant approaches in Organisational Psychology, Industrial Psychology, and Human Resources and summarised their major findings when it comes to tools to measure personality and competences of employees. These attributes are used in team composition processes that we present in the next chapters, i.e. Chapter 4 and Chapter 5.

The most explored schemes to measure personality are using subjective self-assessment questionnaires, i.e. personality tests. We divide the most popular personality schemes used for team composition into two approaches:

- Individual Traits Approach (that includes the Five Factor Model, which uses five broad dimensions to describe human personality; and the Myers-Briggs Type Indicator (MBTI) scheme designed to indicate psychological preferences in how people perceive the world and make decisions), and
- Team Balance Approach (that includes Belbin theory, which provides a theory on how different role types influence teamwork; the Fundamental Interpersonal Relations Orientations (FIRO), which uses three needs to balance teams; and the Post-Jungian Personality Theory, which bases his theory on balancing both personality traits and gender within each team).

Regarding the Individual Approach, it is based on the presumption that some individuals are simply better than others, when it comes to working in teams. The most popular test in this approach is the Five Factor Model (FFM). However, research testing the relationship between FFM and team performance produced mixed results. The Team Balance Approach rather tries to find compatibility between different personalities and based on that, the best team members configurations. Surprisingly, research examining the team balance approach is quite limited. However, the team composition results based on the Post-Jungian Personality theory developed by Douglas J. Wilde seem promising as within a decade this novel approach has tripled the faction of Stanford teams awarded national prizes by the Lincoln Foundation. Therefore, in the next Chapter we use this theory to compose *Congenial Teams*.

When it comes to an assessment of employees' competences, there are various methods that can be used in organisations (i.e. cognitive assessments, work samples, employee achievement history, competence-based interviews and assessment centres, as well as training activity, self-questionnaires and peers-assessments). All these sources can help human resources to measure how competent is their workforce, although all of them require a significant amount of work connected to developing a shared organisational competence model. Therefore, organisations can use Multiple Intelligences Theory to facilitate and quickly obtain some knowledge about employees' competences. In the next Chapter we present an experiment with students, where we used Multiple Intelligences Theory to assess their competences. We use this method as it gives us a global picture of the abilities of students.

Additionally, all assessment methods presented in this Chapter allow for a significant amount of bias. This bias could be diminished if the competences' evaluation process was a constant process in the organisation, where employees (i.e. team peers performing the task together or project managers responsible for those tasks) frequently assess one's another competences based on the tasks performed. Additionally, allowing peers to comment on the assessments of others could discover potential biases. In Chapter 6, we present an algorithm that is able to identify biased peers and lower their importance in the contribution to a final performance rating. Before that, in Chapter 4 we present methods that use employees' individual attributes as an input for team composition processes.

Chapter 4

Synergistic Team Composition Model

4.1 Introduction

In this Chapter, we present a *Team Composition* model for our management organisational workflow presented in Chapter 1 (see figure 1.1).

Teams provide a structure and means of bringing together people with a suitable mix of individual attributes. They can also encourage creativity, the exchange of ideas, facilitate perspective taking, motivate, give job support and actually extend individuals' capabilities. In turn, a suitable team can support real-time problem solving and initiative, improve the overall productivity, and the quality of the performed tasks [Katzenbach and Smith, 2015]. Additionally, teams can lead to a higher job satisfaction. For example, Katzenbach and Smith (2015) observed a specific sense of humor on the job within the top-performing teams as a method to deal with the task pressure.

However, sometimes a team may work less effectively than initially expected. Even teams with comparably competent members can have radically different levels of performance. When we analysed the literature on team composition, formation and teamwork (see Chapter 2 for your reference), we established that one of the crucial findings to maximize team performance in both Organisational Psychology and Computer Science is that team members have to be heterogeneous [Bear and Woolley, 2011; Hanna and Richards, 2015; Horwitz and Horwitz, 2007; Marcolino et al., 2013; Mathieu et al., 2008; Osatuyi, 2012; West, 2012b; Wilde, 2013]. That is, team members should differ in some attributes. It was also found that main factors influencing team performance include competences, experiences, age and gender as well as personality [Arnold and Randall, 2010; Mount et al., 1998; Navarro et al., 2011; Rangapuram et al., 2013; Schutz, 1958; West, 2012b; White, 1984; Wilde, 2009]. Well performing teams bring together complementary competences, experiences and points of view that, by definition, surpass those of any team member by her own.

Given this background, in this chapter, we focus on how to compose a single *synergistic* team for a single complex task based on individuals' competences, personality and gender. In detail, the scenario is as follows. We have a task to be solved that requires a set of competences with given competence levels. We have a pool of human employees with varying genders, personalities, and competence levels. Personality traits of individuals are obtained through the Post-Jungian Personality Theory (see Chapter 3 for details of the questionnaire). Our goal is to compose teams to be both *proficient* (cover the required competences) and *congenial* (balance gender and psychological traits). We define the *synergistic value* of a team as its proficiency degree and balance in terms of personality and gender. We empirically evaluate our team composition model using real data in an education scenario. We show that our model predicts team performance better than experts who know employees' social situation, background and competences.

Outline. The remaining of this chapter is structured as follows. First, in section 4.2 we introduce the basic notions of employee, personality, competence, team, and task type. In Section 4.3 we discuss how to solve a competence assignment problem. Next, in Section 4.4 we define a method to evaluate how proficient (competent given a particular task) and congenial (balanced in personality and gender) a team is and we combine these two notions to determine a *synergistic value* of a team as a weighted sum of proficiency and congeniality. Finally, in Section 4.5 we discuss the experiments that we performed in order to evaluate the effectiveness of our model in predicting the performance of teams. Finally, Section 4.6 discusses the potential applications of our model in organisations.

4.2 Basic Definitions

In our work we consider that each employee is a human. Each employee is characterised by the following attributes:

1. **A unique *identifier***: It distinguishes an employee from others.
2. ***Gender:*** {*male, female*} stands for their gender.
3. **A *personality***: represented by four personality traits, each one within $[-1, 1]$. For example for the Feeling-Thinking (TF) dimension, a value between -1 and 0 means that a person is of feeling type, and a value between 0 and 1 means she is of thinking type.[1]
4. **A *set of competences***: A competence integrates knowledge, skills, personal values, and attitudes that enable an employee to act correctly in a job, task or situation [Roe, 2002a]. Each employee is assumed to possess

[1] In Chapter 3 we discussed details of the Post-Jungian Personality questionnaire to measure personality traits of individuals. Please note that the numerical data collected using the MBTI questionnaire can also be used for this purpose.

a set of competences with associated competence levels. Associated levels of competences can adjust as the employee learns.

Next, we formally define the notions of personality and employee.

Definition 4.1. A *personality profile* is a vector $\langle sn, tf, ei, pj \rangle \in [-1, 1]^4$ of personality traits.

We denote by $C = \{c_1, \ldots, c_m\}$ the whole set of competences, where each element $c_i \in C$ stands for a competence.

Definition 4.2. An *employee* is represented as a tuple $\langle id, g, \mathbf{p}, l \rangle$ such that:

- id is the employee's identifier;
- $g \in \{man, woman\}$ stands for employee gender;
- $\mathbf{p}$ is a personality profile vector;
- $l : C \to [0, 1]$ is a function that assigns the quality level of the outcome with respect to competence c. We will refer to $l(c)$ as the *competence level* of the employee for competence c. We assume that when an employee does not have a competence (or we do not know about it), the level of this competence is zero.

Henceforth, we will note the set of employees as $A = \{a_1, \ldots, a_n\}$. Moreover, we will use super-indexes to refer to employees' components. For instance, given an employee $a \in A$, id^a will refer to the id component of employee a.

Next, we move on to the definition of a team.

Definition 4.3 (Team)**.** A *team* is any subset of A with at least two employees. We denote by $\mathcal{K}_A = (2^A \setminus \{\emptyset\}) \setminus \{\{a_i\} | a_i \in A\}$ the set of all possible teams in A.

Given a team K, we note by $w(K)$ and by $m(K)$ the number of women and men in the team respectively.

We assume that employees in teams coordinate their activities for mutual benefit.

Next, we define a task type and a task. We understand a task as a particular instance of a *task type*. A task type determines the competence levels required to solve the task as well as the importance of each competence with respect to the others. Additionally, task types differ in their character requiring different levels of proficiency. For instance, some task types may require a high level of creativity because they were never performed before (no qualified employees in this matter). Others may require a highly proficient team (as it is the case for rescue teams). Formally, a task type is defined as follows.

Definition 4.4. A task type τ is a tuple $\langle \lambda, \{(c_i, l_i, w_i)\}_{i \in I_\tau} \rangle$ where I_τ is the index set of the required competences and:

- $\lambda \in [0, 1]$ is the importance given to proficiency; the higher the value of λ, the higher the importance for the proficiency of a team.
- $c_i \in C$ is a competence required to perform the task;
- $l_i \in [0, 1]$ is the required competence level for c_i;
- $w_i \in [0, 1]$ is the importance of competence c_i for the success in solving an instance of task type τ; and $\sum_{i \in I_\tau} w_i = 1$.

A task is an instance of a task type that specifies how many employees must be included in a team. We define a task as follows:

Definition 4.5. A *task* t is a tuple $\langle \tau, m \rangle$ such that τ is a task type and m is the required number of employees, where $m \geq 2$.

Henceforth, we denote by T the set of tasks and by $\mathcal{T}$ the set of task types. Moreover, we will note as $C_\tau = \{c_i | i \in I_\tau\}$ the set of competences required by task type τ.

4.3 A competence assignment problem

In this section we discuss how to assign competences of members of a team to competence requirements determined by a task type. First, we define an assignment. Next, we identify some properties of competence assignments that can help us determine if an assignment is appropriate for our task type. Finally, we discuss an assignment as an optimisation problem where we want each competence assigned to at least one employee and each employee assigned to at least one competence so that the total cost (that is both under-proficiency and over-proficiency) of the assignment is minimal with respect to all such assignments.

4.3.1 Defining an assignment

Employees must feel both accountable and useful when working in a team. Hence, each team member must be responsible for at least one competence. This is expressed as a *competence assignment* between competences and employees:

Definition 4.6. Given task type τ and a team $K \in \mathcal{K}_\mathcal{A}$, a competence assignment is a function $\eta : K \to 2^{C_\tau}$ satisfying that $C_\tau = \bigcup_{a \in K} \eta(a)$. We note by Θ_τ^K the set of competence assignments for task type τ and team K.

We define a list of employees assigned to each competence as follows.

Definition 4.7. Given task type τ, team K, and competence assignment η, the set $\delta(c_i) = \{a \in K | c_i \in \eta(a)\}$ stands for those employees responsible of competence c_i.

Now we are ready to discuss the properties of competence assignments.

4.3.2 Properties of competence assignments

In this subsection we identify some of the most important competence assignment properties. Please note that this list of properties is not exhaustive and we can define many more assignment properties that were not listed in here. It is rather supposed to help us understand the logic behind the usage of particular competence assignments rather than give us an ontology of all potential assignment properties.

Complete Assignment We say that an assignment is *complete* when all team members are responsible for all required competences. The formal definition of complete assignment is as follows.

Definition 4.8. An assignment η is *complete* for task type τ and team $K \in \mathcal{K}_{\mathcal{A}}$ iff for all competences $c_i \in C_\tau$ and for all agents $a \in K$ $c_i \in \eta(a)$.

We may want to know if an assignment is complete when we want to measure a general team competence. It is also important to impose an assignment to be complete when we are looking for a team with homogeneous competences. For instance, if our task was to write a program in a certain programming language, we would want all team members to know that particular programming language.

Fully proficient assignment In some task types it is important to make sure that all employees responsible for required competences are proficient enough. In other words, given a team and a task type, we must know if for each competence there is an employee whose competence level of the assigned competence is at least as high as the required level. We refer to this kind of assignment as *fully proficient*. Formally, we define it as follows:

Definition 4.9 (Fully proficient assignment)**.** Given a set of agents A, task type τ and team $K \in \mathcal{K}_{\mathcal{A}}$, assignment η makes team K fully proficient for task type τ iff η is a competence assignment and for all competences $c_i \in C_\tau$ there is $a \in K$ such that $c_i \in \eta(a)$ and $l^a(c_i) \geq l_i$.

This property is especially important when a team has only one chance to succeed (like rescue missions or a complicated health interventions) as failing to perform a task could have serious health or life implications.

Educational Assignment It has been recognized that organisations can only learn if teams in the organisation learn through the sharing of knowledge and experience among employees [Chan et al., 2003]. To support co-learning we might want to assign for each competence the following roles to two different employees in the team:

1. Responsible: an employee that will be accountable for the competence and that must be under-competent given the competence level requirement.
2. Mentor: an employee supervising the work of the responsible employee, that is over-competent.

We refer to this kind of assignment as *educational assignment.* The formal definition of educational assignment goes as follows.

Definition 4.10. An assignment η is *educational* for task type τ and team $K \in \mathcal{K}_{\mathcal{A}}$ iff for all competences $c_i \in C_\tau$, there are $a_1, a_2 \in K$ such that $c_i \in \eta(a_1)$, $c_i \in \eta(a_2)$ and $l_i^{a_1} < l_i \leq l_i^{a_2}$.

Educational assignment is suitable when we want a junior employee to learn from a senior. For instance, performing a peer programming task.

Lavish Assignment For tasks that are time sensitive, it is crucial to act both both quickly and accurately. For this reason, we may want all team members to use their strongest competences to contribute to the team's success. Therefore, we must assign the most proficient employee from a team to each required competence. We refer to this kind of assignment as *lavish assignment.* Formally, we define the lavish assignment as follows.

Definition 4.11. An assignment η is *lavish* for task type τ and team $K \in \mathcal{K}_{\mathcal{A}}$ iff for all competences $c_i \in C_\tau$, if $a \in K$ and $c_i \in \eta(a)$ then $l^a(c_i) \geq l_i$ and $\nexists b \in K$ such that $c_i \notin \eta(b)$ and $l^b(c_i) > l^a(c_i)$.

Exact Assignment We can also compose the minimally competent teams by assigning employees that have minimal levels of required competences. We refer to this kind of assignment as *exact assignment.* The exact assignment might be useful in dynamic scenarios, where we want to "spare" the most proficient employees in case a new, more difficult task arrives. Additionally, performing too simple tasks may cause a drop in motivation [Bashur et al., 2011]. The formal definition of exact assignment is as follows.

Definition 4.12. An assignment η is *exact* for task type τ and team $K \in \mathcal{K}_{\mathcal{A}}$ iff an assignment for all competences $c_i \in C_\tau$, if $a \in K$ and $c_i \in \eta(a)$ then $l_i \leq l_i^a$ and $\nexists b \in K$ such that $c_i \notin \eta(b)$ and $l_i^b < l_i^a$.

Optimal Assignment In various task types we want to know if an assignment of team members to required competences is *optimal.* That is, we want to make sure that for each competence a distance between the required competence level and the actual team member competence level assigned to that competence is minimal. Formally, we define it as follows.

Definition 4.13. An assignment η is *optimal* for task type τ and team $K \in \mathcal{K}_{\mathcal{A}}$ iff for all competences $c_i \in C_\tau$, if $a \in K$ and $c_i \in \eta(a)$ then $\nexists b \in K$ such that $c_i \notin \eta(b)$ and $|l_i^a - l_i| > |l_i^b - l_i|$.

This assignment property is similar to the exact assignment property with the difference that here, we do not require assigned agents to be at least as competent as required by the task type. We rather minimize the total *cost* of the assignment (in terms of under- and over-proficiency). This property is typically used in organisations when we want to make sure a task is performed well but we do not want to overpay for an over-proficient expert.

Inclusive Assignment In various task types we want to know if an assignment of team members to required competences is *inclusive*. That is, we want to make sure that each team member is assigned to at least one competence. Formally, we define it as follows.

Definition 4.14. An assignment η is *inclusive* for task type τ and team $K \in \mathcal{K}_{\mathcal{A}}$ iff for all employees $a \in K$ $|\eta(a)| \geq 1$.

This property is used when we want each team member to be responsible for at east one part of a task. This is particularly important in education, where no one should be excluded from teamwork.

4.3.3 Computing optimal inclusive assignment

There are many different competence assignments that can be defined. Depending on the way we assign competences the quality of team performance will vary. For instance, if a task type requires a team member who speaks Chinese and we have an employee who does but we assign him to another competence, then the team will fail even though it has competences to succeed. Thus, solving a competence assignment problem is one of the most important problems when looking for a well performing team.

In this chapter we treat a competence assignment problem as an optimisation problem. That is, we want to impose an assignment to be both optimal and inclusive, i.e. get each competence assigned to at least one employee and each employee assigned to at least one competence so that the total cost (that is both under-proficiency and over-proficiency) of the assignment is minimal with respect to all such assignments. This assignment problem can be efficiently solved using the minimum cost flow model [Ahuja et al., 1993]. For instance, in [Orlin, 1993], it was proven that the minimum cost flow problem can be solved in $O(m \cdot log(n) \cdot (m + n \cdot log(n)))$ time on a network with n nodes and m arcs.

Formally, let $G = (N, E)$ be a directed network defined by a set N of nodes and a set E of directed arcs. There are four types of nodes: (1) one source node; (2) $|K|$ nodes that represent agents in team K; (3) $|C_\tau|$ competence requests that form task type τ; and (4) one sink node. Each arc $(i, j) \in E$ has an associated cost $p_{ij} \in \mathbb{R}^+$ that denotes the cost per unit flow on that arc. We also associate with each arc $(i, j) \in E$ a capacity $u_{ij} \in \mathbb{R}^+$ that denotes the maximum amount that can flow on the arc. In particular, we have three kinds of edges: (1) Supply arcs connecting the source to agent nodes. Each of these arcs has zero cost and a positive capacity $u_{ij} = \left\lceil \frac{|C_\tau|}{|A|} \right\rceil$ which define how many competences at most can be assigned to each agent. (2) Transportation arcs used to ship supplies. Every transportation arc $(i, j) \in E$ is associated with a cost p_{ij} that is equal to:

$$p_{ij} = \begin{cases} (l^{a_i}(c_j) - l_j) \cdot (1 - \upsilon) \cdot w_j & \text{if } l^{a_i}(c_j - l_j) \geq 0 \\ -(l^{a_i}(c_j) - l_j) \cdot \upsilon \cdot w_j & \text{if } l^{a_i}(c_j - l_j) < 0 \end{cases}$$

where $\upsilon \in [0, 1]$ is the penalty given to the under-proficiency of team K (we explain these notions with more detail later on, in section 4.4.1) and $w_j \in$

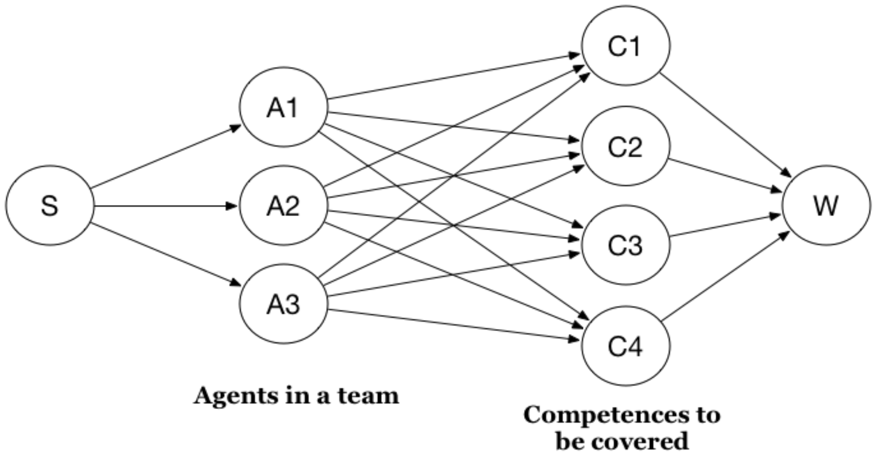


Figure 4.1: An example of an assignment graph

$[0, 1]$ is the importance of competence c_j for the success of task type τ given to particular competence (see definition 4.4). (3) Demand arcs connecting the competence requests nodes to the sink node. These arcs have zero costs and positive capacities u_{ij} which equal the demand for each competence.

Thus, a network is denoted by (G, w, u, b). We associate with each node $i \in N$ an integer number $b(i)$ representing its supply. If $b(n) > 0$ then n is a source node, if $b(n) < 0$ then n is a sink node.

Example Let us consider a team of three agents $K = \{a_1, a_2, a_3\}$:

- $a_1 = \langle id_1, \text{'}woman', p_1, [l(c_1) = 0.9, l(c_2) = 0.5]\rangle$
- $a_2 = \langle id_2, \text{'}man', p_2, [l(c_2) = 0.2, l(c_3) = 0.8]\rangle$
- $a_3 = \langle id_3, \text{'}man', p_3, [l(c_2) = 0.4, l(c_4) = 0.6]\rangle$

and task type τ containing four competence requests $\{(c_1, 0.8, 0.25), (c_2, 0.6, 0.25), (c_3, 0.6, 0.25), (c_4, 0.6, 0.25)\}$. The penalty given to under-proficiency is equal to $\upsilon = 0.6$.

Our goal is to assign agents to competence requests, so that: (1) every agent is responsible for at least one competence, (2) every competence is covered by at least one agent, (3) the overall "cost" in minimal. As shown in figure 4.1, we build a graph out of $n = 9$ nodes that is: one source node (S), three agents nodes $(A_1 - A_3)$, four competences nodes $(C_1 - C_4)$ and a sink node (W). Next, we add edges: (1) between source node S and all agent nodes $A_1 - A_3$ that have a cost $p_{si} = 0$ and capacity $u_{si} = 2$ for all i, as the maximum number of competences assigned to one agent cannot be bigger than two if we want to make sure that all agents are assigned to at least one competence; (2) between agent nodes $(A_1 - A_3)$ and competence nodes $(C_1 - C_4)$, where each capacity $u_{ij} = 1$ and we calculate costs p_{ij} according to the above equation. For instance, the cost between C_1 and C_4 is equal to: $(0.9 - 0.8) \cdot (1 - 0.6) \cdot 0.25 = 0.01$. (3) edges between competence nodes $C_1 - C_4$ and sink node W that have costs $p_{jw} = 0$ and capacities $u_{jw} = 1$ to impose that each competence is assigned. The built graph is passed to a solver (for instance, see an implementation of Goldberg and Tarjan, 1990) to get an optimal assignment. This solver will return an optimal assignment for team K that we denote as η^*.

4.4 Synergistic Team Composition Model

In this section we define a model to evaluate and compose teams. In what follows, we refer to this model as STCM. We start by introducing the notion of *proficiency* as matching between a team and a task given an assignment. Next, we move on to discuss the notion of *congeniality* as a measure the diversity of a team. Finally, we combine these two notions to calculate the *synergistic* value of a team.

4.4.1 Evaluating team proficiency

Given a team and a task, we want to calculate the *degree of proficiency* of the team as a whole, noted u_{prof}. In other words, our aim is to match each competence with the employee(s) whose personal competence level is closer to the task competence level requirement. With this we aim at avoiding both *under-proficient* and *over-proficient* allocations as both of those scenarios are ominous for team performance. In the first case, under-proficient employees may get frustrated because they do not have enough knowledge to cope with the assigned competence requirements. In the second case, *over-proficient* employees may get distracted and unmotivated because of the easiness of a job they are asked to do [Bashur et al., 2011]). We define the degrees of under-proficiency and over-proficiency as the distances between the competence levels required by the task and those offered by the assignment as follows.

Definition 4.15 (Degree of under-proficiency)**.**

Given task type τ, a team K, and a competence assignment η, we define the team's degree of under-proficiency for the task as:

$$u(\eta) = \sum_{i \in I_\tau} w_i \cdot \frac{\sum_{a \in \delta(c_i)} |\min(l^a(c_i) - l_i, 0)|}{|\delta(c_i)| + 1}$$

Definition 4.16 (Degree of over-proficiency)**.**

Given task type τ, a team K, and an assignment η, the team's degree of over-proficiency for the task is defined as:

$$o(\eta) = \sum_{i \in I_\tau} w_i \cdot \frac{\sum_{a \in \delta(c_i)} \max(l^a(c_i) - l_i, 0)}{|\delta(c_i)| + 1}$$

We remind the reader that w_i is the importance of competence c_i for task type success, l_i is the required level of competence c_i, $\sigma(c_i)$ are the agents assigned to c_i and l_i^a is the level of competence c_i of agent a.

Given a competence assignment for a team, we can determine its proficiency degree to perform the task by calculating a weighted average of team's over-proficiency and under-proficiency. The weight may be used to penalize more the team's under-proficiency, as some tasks strictly require teams to be at least as proficient as defined in the task type.

Definition 4.17. Given task type τ, a team K and an assignment η, the proficiency degree of the team to perform the task is defined as:

$$u_{prof}(K) = \max_{\eta \in \Theta_\tau^K} (1 - (\upsilon \cdot u(\eta) + (1 - \upsilon) \cdot o(\eta)) \tag{4.1}$$

where $\upsilon \in [0, 1]$ is the penalty given to the under-proficiency of team K.

Notice that the larger the value of υ the higher the importance of the proficiency degree of team K, while the lower the value υ, the less important its under-proficiency. Therefore, if we want to penalise teams that cannot cope with the competence requirements (i.e. they are under-competent) we need to choose a large value for υ. And similarly a small υ to penalise teams with members clearly over-competent. Although the exact value to choose will depend on the particular task type and the goal for team composition. If the objective is to favour *effective* teams we should penalize more their under-proficiency and thus select a significantly large value for υ. Given these definitions, $u_{prof}(K)$ is correctly defined for any team, task type and competence assignment:

Proposition 1. For any task type τ, team K, and $\eta \in \Theta_\tau^K$, $u(\eta) + o(\eta) \in [0, 1)$ and $0 \leq u_{prof}(K) < 1$.

Proof. Given that (1) $l^a(c_i) \in [0, 1)$ and $l_i \in [0, 1)$; (2) If $\min(l^a(c_i) - l_i, 0) < 0$ then $\max(l^a(c_i) - l_i, 0) = 0$; and (3) If $\max(l^a(c_i) - l_i, 0) > 0$ then $\min(l^a(c_i) - l_i, 0) = 0$. Thus, from (1–3) we have $|\min(l^a(c_i) - l_i, 0)| + \max(l^a(c_i) - l_i, 0) \in [0, 1)$. Let $n = |\{a \in \delta(c_i) | l^a(c_i) - l_i > 0\}|$, then obviously it holds that $\frac{(n+1) \cdot (|\min(l^a(c_i) - l_i, 0)| + \max(l^a(c_i) - l_i, 0))}{n+1} \in [0, 1)$ and as $|\delta(c_i)| \leq (n + 1)$ then $\frac{\sum_{a \in \delta(c_i)} (|\min(l^a(c_i) - l_i, 0)| + \max(l^a(c_i) - l_i, 0))}{n+1} \in [0, 1)$ holds; and since $\sum_{i \in I_\tau} w_i = 1$, $\sum_{i \in I_\tau} w_i \cdot \frac{\sum_{a \in \delta(c_i)} (|\min(l^a(c_i) - l_i, 0)| + \max(l^a(c_i) - l_i, 0))}{n+1}$ $\in [0, 1)$ holds; Finally, this equation is equivalent to: $\sum_{i \in I_\tau} w_i \frac{\sum_{a \in \delta(c_i)} (|\min(l^a(c_i) - l_i, 0)|)}{n+1} + \sum_{i \in I_\tau} w_i \frac{\sum_{a \in \delta(c_i)} (\max(l^a(c_i) - l_i, 0))}{n+1}$ $\in [0, 1)$ which in turn is equivalent to $u(\eta) + o(\eta) \in [0, 1)$. □

Function $u_{prof(K)}$ is used to measure how proficient a team is for a given competence assignment. However, the degree of proficiency alone does not guarantee that the team will succeed at performing it. Therefore, in the next subsection we present an evaluation function to measure *congeniality* within teams. Unlike our measure for proficiency, which is based on considering a particular competence assignment, our congeniality measure will solely rely on the personalities and genders of the members of a team.

4.4.2 Evaluating team congeniality

According to Davey [Davey, 2017], the only truthful collaboration is the one containing tension, disagreement, and conflict as these improve the value of the ideas, expose the risks inherent in plan, and lead to enhanced trust among the team members. Recent studies in organisational psychology show that there is a trade-off between the creative productivity caused by "meta-cognitive conflict" and "harmony" — good feeling — on a team [Bradley and Hebert, 1997]. This conflict is generated by people having different views of the world (associated with opposing personality and gender), whereas harmony comes from agreement between people with similar personalities [Wilde, 2013].

Based on these findings Douglas J. Wilde [Wilde, 2009] compiled heuristics to successfully compose teams. Inspired by his work we construct cognitively diverse teams using the psychological function pairs SN and TF, the psychological attitudes PJ and EI, and gender. In order to mathematically capture those heuristics, we define a novel utility function for *congeniality*, u_{con}.

Inspired by the experiments of Douglass J. Wilde [Wilde, 2009] described in Chapter 3, our congeniality measure follows the following set of rules:

1. It values more teams whose sensing-intuition (SN) and thinking-feeling (TF) personality dimensions are as diverse as possible;
2. It prefers teams with at least one employee with positive EI, TF and PJ dimensions, namely an extrovert, thinking and judging employee (called ETJ personality);
3. It values more teams with at least one employee with negative EI dimension, namely introvert; and
4. It prefers gender balance in a team.

Therefore, the higher the congeniality value of a team, the more diverse the team. Formally, this team utility function is defined as follows:

$$u_{con}(K) = u_{SNTF}(K) + u_{ETJ}(K) + u_I(K) + u_{gender}(K), \tag{4.2}$$

where the different parameters are explained next.

- $u_{SNTF}(K) = \sigma(K, SN) \cdot \sigma(K, TF)$ measures the diversity in a team, where $\sigma(K, SN)$ and $\sigma(K, TF)$ stand for the standard deviations over the SN and TF personality traits of the members of team K. The larger the values of σ_{SN} and σ_{TF} the larger their product will be, and hence the larger the personality diversity along the SN and TF dimensions within a team.
- $u_{ETJ}(K) = \max_{a \in K^{ETJ}}[\max(\boldsymbol{\alpha} \cdot \mathbf{p}, 0), 0]$ measures the utility of counting on ETJ personalities, being $K^{ETJ} = \{a \in K | tf^a > 0, ei^a > 0, pj^a > 0\}$ the set of employees exhibiting ETJ personality, $\boldsymbol{\alpha} = (0, \alpha, \alpha, \alpha)$ is a vector, and α is the importance of counting on an extrovert, thinking, and judging employee (ETJ personality). The maximum variance of any distribution

over an interval $[a, b]$ corresponds to a distribution with the elements evenly situated at the extremes of the interval. The variance will always be $\sigma^2 \leq ((b-a)/2)^2$. In our case with $b = 1$ and $a = -1$ we have $\sigma \leq 1$. Then, to make the four factors equally important and given that the maximum value for $\mathbf{p^i}$ (the personality profile of employee a_i) would be $(1, 1, 1, 1)$ a maximum value for α would be $3\alpha = ((1-(-1))/2)^2 = 1$, as we have the factor $\sigma_{SN} \cdot \sigma_{TF}$, so $\alpha \leq 0.33(3)$. For values situated in the middle of the interval the variance will be $\sigma^2 \leq \frac{(b-a)^2}{12}$, hence a reasonable value for α would be $\alpha = \frac{\sqrt{(1-(-1))^2)/12}}{3} = 0.19$.

- $u_I(K) = \max_{a \in K}[\max(\boldsymbol{\beta} \cdot \mathbf{p}, 0), 0]$ is the utility of counting on an introvert employee, where $\boldsymbol{\beta} = (0, 0, -\beta, 0)$ is a vector and β is the importance of introvert employees. A similar reasoning to (2) shows that $\beta \leq 1$.
- $u_{gender}(K) = \gamma \cdot \sin(\pi \cdot g(K))$ measures the importance of gender balance, where γ is a parameter to weigh the importance of gender balance, and $g(K) = \frac{w(K)}{w(K)+m(K)}$ calculates the ratio of women in a team ($w(K)$ and $m(K)$ are functions counting the number of women and men, respectively). A team K is perfectly gender-balanced iff $w(K) = m(K)$, and hence $\sin(\pi \cdot g(K)) = 1$. In order to make gender factor as important as the others in the equation we analytically assessed that $\gamma = 0.1$ is a good compromise.

4.4.3 Evaluating synergistic teams

A team K is effective solving a task when it is both *proficient* (covers the required competences) and *congenial* (balances gender and psychological traits so that employees work well together). We obtain its *synergistic value* as a weighted, linear combination of its proficiency (in equation 4.1) and congeniality (in equation 4.2) values as follows:

Definition 4.18. Given a team K, the synergistic value of team K is defined as:

$$s(K) = \lambda \cdot u_{prof}(K) + (1 - \lambda) \cdot u_{con}(K) \tag{4.3}$$

where $\lambda \in [0, 1]$ is the relative importance of the proficiency of team K.

In general, the higher the value of λ, the higher the importance for the proficiency of a team. The setting of the value of λ depends on the task type. For instance, task types that are difficult and performed for the first time (no experts on that matter) require a high level of creativity and exchange of ideas, and hence, congeniality should be more important than proficiency ($\lambda < 0.5$). However, for tasks where team members need to act fast (such as sport competitions or rescue teams) it is crucial for a team to be proficient ($\lambda > 0.5$). For creative task types that require certain levels of both proficiency and congeniality (such as creating a webpage) the value of λ should be set to 0.5 (so that congeniality and proficiency are equally important).

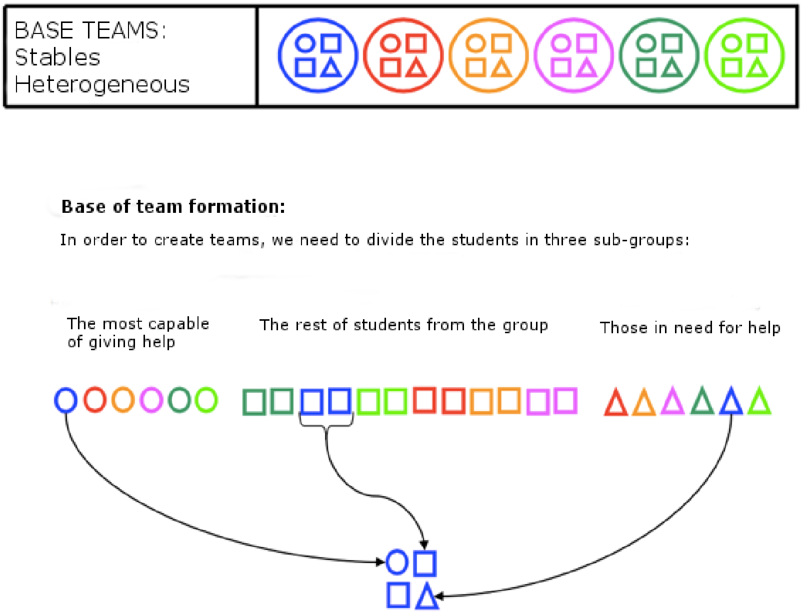


Figure 4.2: Current practice on team composition (from [Amigó, 2018]).

4.5 Experimental Results

In this section we discuss the experiment that we performed in order to pitch our synergistic team composition model (STCM) with the team composition performed by experts (secondary school teachers). This is our very first experiment performed with the purpose of validating the model and finding the best λ value for creative tasks in education. In chapter 5 we discuss further experiments that show the effectiveness of the STCM model.

In what follows, we compare the STCM model with the *teacher model* (see subsection 4.5.1) in terms of how well each of them predicts team performance. Since we observe that the STCM outperforms experts at predicting team performance, we argue that it is the method of choice in the classroom.

During the experiments, since they are performed in the context of education, we refer to employees (defined in Section 4.2) as "students".

4.5.1 Teacher Method

Before discussing the details of our experiment, we start from describing the current practice used in education in Catalonia. In the remaining of this thesis, we refer to this method as the *Teacher method*.

Students learn best when they are actively engaged in the processing of information [Vosniadou, 2003]. One way to involve students in active learning is to have them learn from one another within teams. Research shows that students working in teams tend to learn more and retain the knowledge longer than when the same content is presented by means of other instructional formats; they also appear more satisfied with their classes [Barkley et al., 2014]. However, not just any team promotes learning. In order for learning to be productive, all teams in the classroom should be small (2–4 students) and heterogeneous, that is, representative of the diversity of the whole class and balanced in size. Also, effective education must balance performance across teams, that is, performance should be as homogeneous as possible in the classroom: *No one should be left behind.*

The learning procedures based on the organisation of individuals into small heterogeneous teams is often referred as co-operative learning. It is based on:

- The strength of interpersonal relationships,
- The effectiveness of socialization and integration values,
- The theories on learning steaming from disparity and sociocognitive conflict.

To build teams, teachers currently distribute the students of a class into three rough sub groups (see Figure 4.2):

- students who are capable of helping others,
- students that are in need for help, and
- the rest of students from the group.

To distribute students teachers rely on their knowledge of students, as not only good grades have to be taken into consideration, also personality traits are important. For instance, a student who has very good grades but is lacking teamwork skills will not be included in the first group, and a disruptive student with low grades but with a good disposition to work on themes that really matter to him/her and/or with a strong leadership, can instead be included in the first group.

Each team should have one student of each sub-group, with only the sub-group "the rest of students" allowed to have two students in a team.

4.5.2 Experimental Setting

In this subsection we discuss the settings and the context of our experiment.

- **Place of the experiment:** "Institut Torras i Bages", a state school in L'Hospitalet de Llobregat, Catalonia. It has 500 students in ages varying from 11 to 18. Collaborative work has been implemented in this school for the last 7 years with a steady and significant increase in the scores and quality of the final product that students are asked to deliver.

- **Time of the experiment:** The experiment took five days, it was performed in June 2016.
- **Student and team data:** The experiment was performed upon two groups of students, that is: '3r ESO A' (24 students), and '3r ESO C' (24 students).
- **Measuring Personality:** Using computers and/or mobile phones, students answered the Post-Jungian Personality questionnaire (described in section 3.1, we provide the full test in Appendix B). The screenshot of the application used to collect personalities is shown in figure 4.3. Before answering the questions students were requested to:
  - focus on their inner self,
  - answer truthfully,
  - answer quickly with minimum over-thought,
  - answer individually (not checking with friends), and
  - keep their answers private.

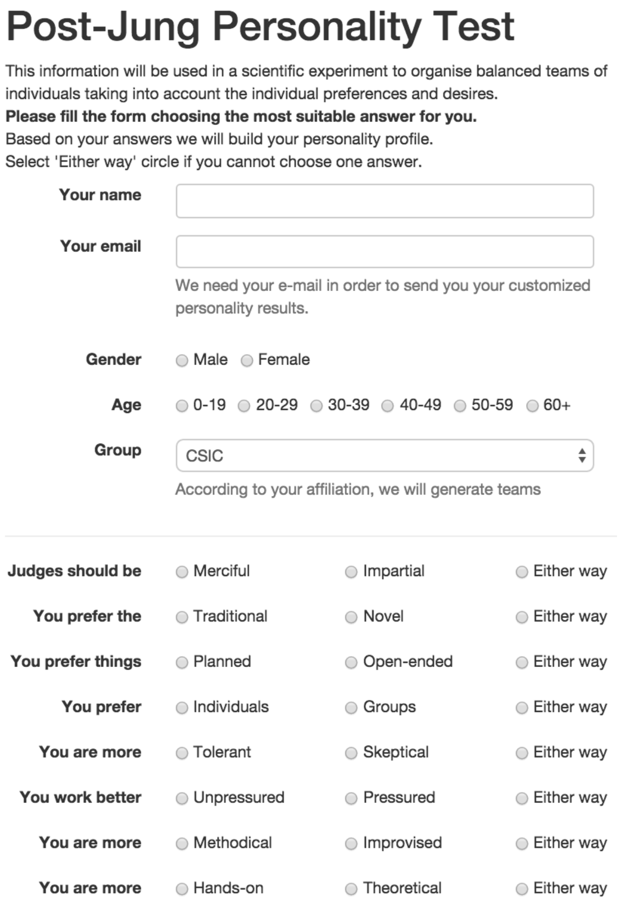
  Students knew that the purpose of the questions was to generate heterogeneous teams, understood the task, and filled in the test as requested.

- **Measuring Competences:** In order to measure competences we used a matrix, provided by tutors, relating each subject to the *intelligences* (competences in the educational context) required for it. For the detailed information on the Multiple Intelligences Theory, please see Section 3.2. The matrix looks as follows:

$$\begin{bmatrix} Catalan \\ Spanish \\ English \\ Nature \\ Physics\ and\ Chemistry \\ Social\ Science \\ Math \\ Physical\ Education \\ Plastic\ Arts \\ Technology \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

  Each column of the matrix represents different Intelligence, respectively: Naturalist, Interpersonal, Logical/Mathematical, Visual/Spatial, Body/Kinaesthetic, Musical, Intrapersonal and Linguistic.

  Based on this matrix we calculated values of intelligences for every student by averaging all her final marks obtained for subjects relevant for this intelligence. For instance, for Body/Kinaesthetic intelligence, we calculate an



Post-Jung Personality Test

This information will be used in a scientific experiment to organise balanced teams of individuals taking into account the individual preferences and desires.
Please fill the form choosing the most suitable answer for you.
Based on your answers we will build your personality profile.
Select 'Either way' circle if you cannot choose one answer.

Your name

Your email

We need your e-mail in order to send you your customized personality results.

Gender Male Female

Age 0-19 20-29 30-39 40-49 50-59 60+

Group CSIC

According to your affiliation, we will generate teams

Judges should be	Merciful	Impartial	Either way
You prefer the	Traditional	Novel	Either way
You prefer things	Planned	Open-ended	Either way
You prefer	Individuals	Groups	Either way
You are more	Tolerant	Skeptical	Either way
You work better	Unpressured	Pressured	Either way
You are more	Methodical	Improvised	Either way
You are more	Hands-on	Theoretical	Either way

Figure 4.3: Snapshot of the web questionnaire used by students to obtain their personality profile.

average of student marks obtained in Nature, Physical Education, Plastic Arts and Technology.

- **Task type:** Students were asked to undertake a set of interdisciplinary activities ("Treball de Sintesi"), which is an obligatory exam performed at the end of each year of the secondary education curriculum in Catalonia, Spain. This assignment takes one week and is designed to check if students have achieved, and to what extent, the objectives set in the various curricular areas. It is a work that encourages teamwork, research, and tests relationships with the environment. In detail, the final assignment that students needed to perform was to plan a restaurant with all the activities connected to it (the market research, the location, the interior design, logo, music selection, budget, menu design, etc). Formally, the task type

$\{(c_i, l_i, w_i)\}_{i\in[1,8]}$ had eight equally important competences, $w_i = 1/8$, with a maximally competence level requirement, $l_i = 1$ as recommended by experts.

- **Team size:** We divided each classroom into teams of size three.
- **Performance evaluation**: Students worked in teams and at the end of every activity presented their work in front of a panel of three teachers that assessed the content, presentation and cooperation between team members (using a standardized rubric on a scale between 1 and 10).
- **STCM settings:** We used the following values for the model:
  - **Evaluating Team Proficiency:** Following the theory on co-learning described in Subsection 4.5.1, we wanted all teams to work similarly well. In other words, we wanted to avoid teams that are particularly strong and others that are particularly weak. Therefore, we decided to penalise teams with over-proficient students as much as teams with under-proficient students. Thus, $\upsilon = 0.5$.
  - **Evaluating Team Congeniality:** We analytically assessed that in order to balance personality requirements (make them equally relevant), the importance given to the different components of the team congeniality function should be as follows: (1) $\alpha = 0.19$, (2) $\beta = 3 \cdot \alpha$, (3) $\gamma = 0.1$.
  - **Assignments:** We solve an assignment problem using the minimum cost flow model and the implementation of [Goldberg and Tarjan, 1990] provided in the ort-tools.[2]

4.5.3 The procedure

The study was performed as follows:

1. We divided each classroom into teams of size three using SynTeam algorithm described further on, in Chapter 5, based on personality (STCM model with $\lambda = 0$). It is an algorithm to partition a classroom of students using the STCM model presented in this Chapter. That is, SynTeam composes a set of teams of even size that are competent as well as personality and gender balanced;
2. Tutors evaluated each team with a mark within $[1, 10]$ representing the expected performance of the team. Note that tutors worked with students for the last 3 years and knew not only the psychological profile of every student from practice, but also the students' social and cognitive capabilities;

[2] https://github.com/google/or-tools/blob/master/src/graph/min_cost_flow.h

3. Based on all students' competences (Intelligences), personalities and actual performances, we calculated synergistic values for different proficiency importance values λ;

4. The teams performed "Treball de Síntesi" and the final marks of students were collected. Having this data, we generated several team performance rankings through different individual team assessment methods:

Rank1 Teacher ranking, generated using the tutors evaluations of teams (meaning the expected performance of the teams);

Rank2 Eleven synergistic values rankings with the varying importance of proficiency $\lambda = \{0, 0.1, \ldots, 1\}$;

Rank3 Actual team performance ranking, i.e. exam marks. This is the *base ranking* to compare against.

Notice that all methods can generate orderings with ties, namely *partial rankings*.

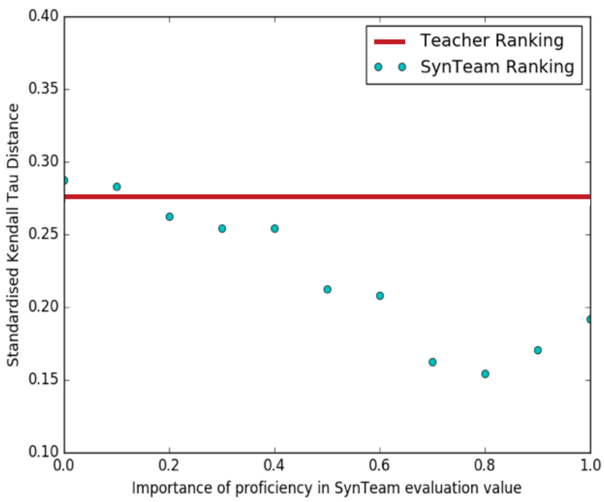


Figure 4.4: Comparison of Generalized Kendall-Tau distances between Teacher ranking and actual ranking versus STCM and actual ranking. The lower the value, the closer to the true ranking (thus, the better the prediction).

4.5.4 Experimental Results

By performing this experiment, we wanted to observe two things. First, we wanted to check the efficiency in the method, that is if our model can predict the performance better than experts. Second, we wanted to know how the rankings were changing when increasing the importance of competences. This was a good method to assess the best value of λ for further experiments.

We compared the teachers' (Rank1) and STCM (Rank2) rankings with the actual performance ranking (Rank3) using the generalised normalized Kendall Tau distance (see Section 2.1 for the background of this method).

The results of the comparison are shown in Figure 4.4. Note that the lower the value of Kendall Tau, the more similar two rankings (and the better the prediction). We observe that the STCM ranking improves as the importance of competences increases. Since the Kendall Tau distance for the teachers' ranking is 0.28, STCM clearly outperforms teachers when competences are included. Notice that there is a gain of close to 50% in prediction accuracy with respect to teachers when the importance of proficiency λ is equal to 0.8. We also calculate the values of Kendall Tau for random (0.42) and reversed (0.9), which shows that both teachers and STCM are better at predicting students' performance than the random method.

4.6 Discussion

In this chapter we proposed a computational model to evaluate teams based on competences, personality and gender of team members. The model is potentially useful for any organisation that faces the need to evaluate their problem solving teams (e.g. a classroom, a company, a research unit). We tested our model in the context of an education with two different classrooms of students. Besides obvious advantages of observing students work in person, this scenario gave us an opportunity to compare our results with real-life, currently used practice. Our results show that our algorithm outperforms experts at predicting teams' performances despite their knowledge about students' social situation, background and competences. In detail, given a task requiring eight competences that are equally important and when using the value of $\alpha = 0.8$ for the synergistic model, we observed that the prediction of team performance is nearly twice better than the prediction done by experts who know the students, their social background, competences, and cognitive capabilities. This shows that our model based on individuals' competences, personality and gender altogether is a good predictor of team performance. In the next chapter, we present more results showing the effectiveness of our synergistic model that are in line with the experiment presented in this chapter.

Chapter 5

Organisational Team Engineering

In this Chapter, we discuss further algorithms for *Team Composition* as a part of our management organisational workflow presented in Chapter 1 (shown in figure 1.1). In the previous Chapter, we introduced and tested the Synergistic Team Composition Model to check the predictivity of our model.

In this chapter we want to address the following situation in organisations: there are multiple *complex tasks that are instances of the same task type*. They have to be solved by different teams of employees of even size. Each task requires that the composed team has at least one employee that shows a minimum level of competence for a given set of competences. We have a pool of employees with varying genders, personalities, and competence levels. The problem is how to partition employees into synergistic teams. That is, teams that are balanced in size, competences, personality, and gender. This is a common scenario in organisations. For instance, in product departments we might need several teams working on adapting the same product to different clients; software teams may work on a product development in parallel; in departments such as sales or customer service it is typical to have tasks of the same task type (such as handling the client's request); finally, it is useful for any task whose purpose is team co-learning.

This chapter makes the following contributions. First, we identify and formalise a new type of real-world problem: the synergistic team composition problem (STCP), requiring *balanced* solutions in terms of team size and team value. Second, we provide an algorithm to optimally solve the STCP. Third, we observe that it is only effective for small instances, hence we develop an algorithm to approximately solve the STCP, when the usage of the optimal solver is too costly. Fourth, we provide a computational comparison of both algorithms over realistic settings in an education context with respect to time and quality that shows a quality-time trade-off. And fifth, we empirically evaluate the approximate algorithm using real data. Our statistically significant results show that our algorithm outperforms experts at composing well-performing teams despite their knowledge about students' social situation, background and competences.

Outline. The remaining of this Chapter is structured as follows. Section 5.1 presents the Synergistic Team Composition Problem (STCP). Section 5.2 discusses optimal and approximate algorithms to solve the synergistic team composition problem. Then, Section 5.3 presents computational comparison of both algorithms. Next, Section 5.4 discusses results of our algorithm in the context of team composition in the classroom. Finally, Section 5.5 discusses our approach and future work.

5.1 The Synergistic Team Composition Problem

In this section we formalise the problem that we want to solve in this Chapter.

Given a set of employees A, our goal is to split them into teams of even size so that each team, and the whole partition of employees into teams, is balanced in terms of competences, personality and gender. We shall refer to these balanced teams as *synergistic teams*, meaning that they are both congenial and proficient.

Therefore, we can regard our team composition problem as a particular type of set partition problem. We will refer to any partition of A as a team partition. However, we are interested in a particular type of team partitions, namely those where teams are constrained by size m as follows.

Definition 5.1. Given a set of employees A, we say that a team partition P_m of A is constrained by size m, $|A| \geq m \geq 2$, iff for every team $K \in P_m$, $m \leq |K| \leq m+1$ holds.

As $|K|/m$ is not necessarily a natural number, we may need to allow for some flexibility in team size within a partition. This is why we introduced above the condition $m \leq |K| \leq m+1$. In practical terms, in a partition we want to have teams of sizes differing by at most one employee. Henceforth, we will focus on team partitions constrained by some size. We note by $\mathcal{P}_m(A)$ the set of all team partitions of A constrained by size m.

The question is: which partition to choose? As discussed before, having one excellent team is not enough, we want all teams to be as good as possible (i.e., we want to avoid partitions where some teams perform very well and some very bad). In other words, we want to have teams that show a homogeneous behaviour so that there are no big differences in performance. To do that, we use the synergistic value of a team K, noted as $s(K)$ (presented formally in subsection 4.4.3), as an expectation of its performance. Second, we define the overall performance of a partition as the Bernoulli-Nash product of the individual team synergistic values, as this function gives larger values to homogeneous, i.e., "fair", solutions [Nash, 1950], than other functions like e.g. the sum.

Definition 1. Given a team partition P_m, the synergistic value of P_m is

$$S(P_m) = \prod_{K \in P_m} s(K). \tag{5.1}$$

Given this definition, the STCP is solved by finding the partition with the largest synergistic value. That is, the synergistic team composition problem is cast as the following optimisation problem:

Definition 2. Given a set of employees A the *synergistic team composition problem (STCP)* is the problem of finding a team partition constrained by size m, $P^*_m \in \mathcal{P}_m(A)$, that maximises $S(P_m)$, namely:

$$P^*_m = \underset{P_m \in \mathcal{P}_m(A)}{\arg\max}\ S(P_m)$$

5.1.1 Relation to the coalition formation literature

The STCP is a particular case of a coalition generation problem [Rahwan et al., 2015]. Unfortunately, we cannot benefit from the algorithms in the literature.

In particular, following [Rahwan et al., 2011b], given a STCP we can identify a constrained coalition formation (CCF) game $\mathcal{G} = \langle A, \mathcal{P}_m(A), s \rangle$, where A is the set of employees, $\mathcal{P}_m(A)$ is the set of feasible coalition structures (i.e. team partitions constrained by size m as per definition 5.1), and s is the characteristic function (synergistic value function) that assigns a real value to every coalition (team) that appears in some feasible coalition structure (team partition).

Given the former CCF game, solving the STCP amounts to finding a coalition structure (team partition) with the highest total value. More precisely, the STCP poses a particular type of CCF game, a so-called *basic* CCF game [Rahwan et al., 2015]. Intuitively, the constraints in a basic CCF game are expressed in the form of: (1) sizes of coalitions that are allowed to form; and (2) subsets of employees whose presence in any coalition is viewed as desirable/prohibited. On the one hand, a STCP naturally defines constraints on the size of coalitions. On the other hand, in order to express a STCP as an CCF problem we have to express one positive constraint for each feasible team (i.e., q positive constrains), while the set of negative constraints is empty. The number of positive constraints is so large for the problems we want to solve (i.e. > 3000) that these problems are prohibitive for the algorithm in [Rahwan et al., 2011b].

5.2 Solving the STCP

In this section we detail two different algorithms to solve the synergistic team composition problem (STCP) described above. Before that, we describe how to split employees into a partition (see subsection 5.2.1). In subsection 5.2.2 we discuss how to linearise our problem to solve it using an integer linear programming solver. In subsection 5.2.3, we explain an algorithm to optimally solve the STCP. In subsection 5.2.4, we explain *SynTeam*, which is a heuristic that quickly generates a first solution, and subsequently aims to improve the current solution by applying a randomized local search procedure.

5.2.1 Partitioning the set of employees

We denote by $n = |A|$ the number of employees in A and by b the total number of teams, $b = \lfloor n/m \rfloor$. Note that depending on the cardinality of A and the desired team size, the number of employees in each team may vary in size. For instance, if there are $n = 7$ employees in A and we want to compose duets, we split employees into two duets and one triplet. In general, we define the quantity distribution of employees in teams of a partition, noted $Q : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N} \cup (\mathbb{N} \times \mathbb{N})^2$ as:

$$Q(n,m) = \begin{cases} \{(b,m)\} & \text{if } n \geq m \text{ and } n \bmod m = 0 \\ \{(n \bmod m, m+1), & \\ (b - (n \bmod m), m)\} & \text{if } n \geq m \text{ and } n \bmod m \leq b \\ \{(0,m)\} & \text{otherwise} \end{cases} \tag{5.2}$$

Hence, $Q(n,m)$ is the quantity distribution of employees in teams of sizes m and $m+1$; these are called *feasible* teams. Beyond these cases, there is no way to compute a partition constrained by m (see def. 5.1). If so, $m' \leq m$, $m' = \lfloor n/(b+1) \rfloor$ is the largest value smaller than m that can be used to compute partitions.

5.2.2 Linearising the STCP

Notice that the STCP poses a non-linear optimisation problem. In what follows, we discuss how to linearise the STCP so that it can be optimally solved by an integer linear programming solver.

Given task $t = \langle \tau, m \rangle$ and a set of employees A, and the quantity distribution of employees $Q(n,m)$, the number of feasible teams is $q = \binom{n}{m} + \min(n \bmod m, 1) \cdot \binom{n}{m+1}$. Therefore, let $K_1, \ldots, K_q$ denote the feasible teams in A, and $s(K_1), \ldots, s(K_q)$ their synergistic values concerning task t. Moreover, let b be the number of teams required to form a team partition. Finally, let C be a matrix of size $n \times q$ such that c_{ij} takes on value 1 if student a_i is part of team K_j, and 0 otherwise.

We shall consider the set of binary decision variables x_j, $1 \leq j \leq q$, to indicate whether team K_j is selected or not as part of the optimal solution of the STCP. Then, solving the STCP amounts to solving the following non-linear problem:

$$\max \prod_{j=1}^{q} s(K_j)^{x_j} \tag{5.3}$$

subject to:

$$\sum_{j=1}^{q} x_j = b \tag{5.4}$$

$$\sum_{j=1}^{b} c_{ij} \cdot x_j = 1 \quad \forall 1 \leq i \leq n \tag{5.5}$$

$$x_j \in \{0,1\} \quad 1 \leq j \leq q \tag{5.6}$$

Notice that constraint 5.4 enforces that the number of teams in the team partition is b, whereas constraint 5.5 enforces that the selected teams form a partition by imposing that no student can belong to two selected teams at the same time. Now observe that equation 5.3 —the objective function— is non-linear. Nevertheless, it can be readily linearised if we consider the logarithm of $\prod_{j=1}^{q} s(K_j)^{x_j}$ as our objective function to maximise. Thus, solving the non-linear problem above is equivalent to solving the following binary linear program:

$$\max \sum_{j=1}^{q} x_j \cdot log(s(K_j)) \tag{5.7}$$

subject to: equations 5.4, 5.5, and 5.6.

Solving this linear problem is an extremely complex task. If $b \cdot m = n$ then a set of n elements can be partitioned into b unordered subsets of m elements each, in the following number of ways [Rolf, 2008]:

$$\frac{1}{b!}\binom{n}{m, \cdots, m} = \frac{n!}{b! \cdot (m!)^b} \tag{5.8}$$

Consider now the case of different team sizes in each partition. That is, if $n \geq m$ *and* $n \bmod m \leq b$, then, as given by equation 5.2, there are $(b - (n \bmod m))$ subsets of m elements each and $(n \bmod m)$ subsets of $(m + 1)$ elements each $(b - (n \bmod m) \cdot m + (n \bmod m) \cdot (m + 1) = n)$. The number of unordered partitions in this case is given by the following expression [Rolf, 2008]:

$$\begin{aligned} &\frac{1}{(b-(n \bmod m))! \cdot (n \bmod m)!}\binom{n}{m, \cdots, m, (m+1), \cdots, (m+1)} = \\ &= \frac{n!}{(b-(n \bmod m))! \cdot (n \bmod m)! \cdot (m!)^{(b-(n \bmod m))} \cdot ((m+1)!)^{(n \bmod m)}} \end{aligned} \tag{5.9}$$

5.2.3 An algorithm to optimally solve the STCP

In this subsection we present an algorithm to optimally solve the Synergistic Team composition Problem. The pseudocode of the optimal solver is shown in Algorithm 1. It starts by generating the input for an integer linear programming solver (lines 2 to 5). Line 2 generates all the possible teams of size m as dictated by the quantity distribution $Q(|A|, m)$. Thereafter, lines 3 and 4 compute the best synergistic value per team. That is, these lines compute (1) the competence assignment with the highest proficiency value. This amounts to solving an optimisation problem, as discussed at the end of subsection 4.4.1, and (2) the team's congenial value from the personalities and genders of the team members. Once all synergistic values are computed, we can generate an integer linear programming encoding of the problem like in equation 5.7 (line 5). The generated integer linear program (ILP) can be solved with the aid of an ILP solver (line 6) such as, for instance, CPLEX, Gurobi, or GLPK. Finally, the algorithm returns the team partition together with the competence assignments (line 7).

Algorithm 1 STCPSolver

Require: A ▷ The set of students
Require: $t = \langle \tau, m \rangle$ ▷ Task
Ensure: $(P, \boldsymbol{\eta^*})$ ▷ Best partition found and best assignments
1: $P \leftarrow \emptyset$
2: $[K_1, \dots, K_q] \leftarrow \mathit{GenerateTeams}(A, Q(|A|, m))$
3: **for** $i \in [1..q]$ **do**
4: $\quad (s(K_i), \eta_i^*) \leftarrow \mathit{getBestSynergisticValue}(K_i, t)$
5: **end for**
6: $ILP \leftarrow \mathit{generateILP}([K_1, \dots, K_q], [s(K_1), \dots, s(K_q)], b)$
7: $P \leftarrow \mathit{solve}(ILP)$
8: **return** $(P, \{\eta_i^*\}_{K_i \in P})$

The cost of optimally solving an STCP can be split into: the cost of generating the ILP model, and the cost of solving it. As to the first cost, this comes from: (i) generating all the teams of sizes given by $Q(n, m)$ (line 2); (ii) computing the synergistic values of all teams (lines 3 and 4); (iii) generating a linear programming encoding (line 5). The cost of generating all teams is linear with the total number of teams, and hence $O(q)$. Note that the total number of teams grows rapidly with increasing m and n. Moreover, the cost of computing the synergistic value for each team requires finding the optimal competence assignment. As discussed in Sec. 4.4.1, this can be cast as a minimum cost flow problem and solved in $O(m \cdot log(e) \cdot (m + e \cdot log(e)))$ time, where $e = m \cdot |C_\tau|$, being $|C_\tau|$ the number of competences required by task type τ. Thus, generating the input to an ILP solver becomes increasingly costly as the number of students per team grows.

5.2.4 An approximate algorithm to solve the STCP

Next we present *SynTeam*, an approximate algorithm to solve the synergistic team composition problem described above. *SynTeam* quickly finds an initial, random partition, to subsequently improve it by performing employee swaps, hoping to reach a global optimum. Algorithm 2, which shows the SynTeam's pseudocode, is divided into two phases:

1. ***Find a first team partition***. This part of the algorithm simply builds a partition by randomly assigning employees to teams of particular team sizes. This part goes as follows. Given a list of employees A, we start by shuffling the list so that the order of employees in the list is random (line 1). Next, we determine the quantitative distribution of individuals among teams of size m using function $Q(|A|, m)$ as defined in section 5.2.1 (line 2). For each number of teams (line 4), we define a temporary set *team* to store a current team (line 5). We add to *team* subsequent *size* employees from the shuffled list of employees (line 8). We add the newly created team to the team partition P_{best} that we intend to build (line 11). When reaching line 14, P_{best} will contain a first disjoint subset of teams (a team partition).

Algorithm 2 SynTeam

Require: A ▷ The list of agents
Require: $Q(|A|, m)$ ▷ Quantitative team distribution
Require: n_r ▷ Maximum number of runs without improvement
Require: $n_l \leq n_r$ ▷ Frequency of local search
Require: $P_{best} = \emptyset$ ▷ Initialize best partition
Ensure: $(P, \boldsymbol{\eta})$ ▷ Best partition found and best assignments
$random.shuffle(A)$
if $Q(|A|, m) \neq (0, m)$ **then**
$index = 0$ ▷ Used to iterate over the agent list
for all $(numberOfTeams, size) \in Q(|A|, m)$ **do**
for $l = 1, \ldots, numberOfTeams$ **do**
$team = \emptyset$
for $i \in (0, \ldots, (size - 1))$ **do**
$team = team \cup A[index]$,
$index = index + 1$
end for
$P_{best} = P_{best} \cup \{team\}$
end for
end for
$\boldsymbol{\eta_{best}} = assign\_agents(P_{best})$ ▷ see Subsection 4.3.3
$(c_r, c_l) = (0, 0)$
while $c_r < n_r$ **do**
$(K_1, K_2) = selectRandomTeams(P_{best})$
$(\eta_1^*, \eta_2^*) = assign\_agents(\{K_1, K_2\})$
$contrValue = u(\{K_1, K_2\}, (\eta_1^*, \eta_2^*))$
$(P_{bestCandidate}, bestCandidatevalue) = (\emptyset, 0)$
for all $P_{candidate} \in P_m(K_1 \cup K_2) \setminus \{K_1, K_2\}$ **do**
$(\eta_1^*, \eta_2^*) = assign\_agents(P_{candidate})$
$candidateValue = u(P_{candidate}, (\eta_1^*, \eta_2^*))$
if $candidateValue > bestCandidateValue$ **then**
$P_{bestCandidate} = P_{candidate}$,
$bestCandidateValue = candidateValue$
end if
end for
if $bestCandidateValue > contrValue$ **then**
$P_{best} = replace(\{K_1, K_2\}, P_{bestCandidate}, P_{best})$,
$c_r = 0$
else
$c_r = c_r + 1,\ c_l = c_l + 1$
end if
if $c_l == n_l$ **then**
$P_{best} = local\_search(P_{best})$,
$c_l = 0$
end if
end while
$return(P_{best}, assign\_agents(P_{best}))$
end if

2. ***Improve the current best team partition***. The second phase of the algorithm consists in iteratively improving the current best team partition. The idea is to obtain better team partitions by (1) performing crossovers of two randomly selected teams and by (2) swapping two employees from different teams (which is a kind of *local search*). In detail, the second phase works as follows. First, two teams, K_1 and K_2, are randomly selected from the current team partition (line 17). Then, all possible ways of re-distributing the employees from $K_1 \cup K_2$ into two new teams are evaluated, and the best candidate team partition $K_1 \cup K_2$, called $P_{bestCandidate}$, is selected (lines 21 to 27). If the product of the synergistic utilities of the best candidate teams is larger than the one of K_1 and K_2, teams K_1 and K_2 are replaced by the teams in the best candidate team partition (line 29) and the no-improvement-counter c_r is reset to zero (lines 29 to 31). Otherwise, the values of the no-improvement counter (c_r) and the counter that controls the local search frequency (c_l) are incremented (line 33). If c_l reaches the value of n_l, the local search is initiated (lines 35 to 38). This local search, as we detail below, consists of swapping two single employees from different teams to yield a better team partition. This second phase of the algorithm continues until the counter c_r reaches the maximum number of runs without improvement n_r (line 16). Finally, the best found solution P_{best}, together with the corresponding assignment η for each team, are returned (line 40).

Algorithm 3 *local_search*

Require: P_{best} ▷ Current best partition
Require: b ▷ The total number of teams
Ensure: P ▷ Improved partition
for all $k \in (0, \ldots, b-1)$ **do**
  for all $j \in (0, \ldots, k)$ **do**
    for all $l \in (k+1, \ldots, b)$ **do**
      for all $i \in (0, \ldots, l)$ **do**
        $P = swap(P_{best}, <k, j>, <l, i>)$
        $(\eta^*_{11}, \eta^*_{12}) = assign\_agents(P_{best}[k], P_{best}[l])$
        $(\eta_{21^*}, \eta^*_{22}) = assign\_agents(P[k], P[l])$
        if $u(P(\eta_{11}, \eta_{12})) > u(P_{best}(\eta^*_{21}, \eta^*_{22}))$ **then**
          $return(P)$
        end if
      end for
    end for
  end for
end for
$return(P_{best})$

Algorithm 3 shows the pseudo-code of the low-cost local search used in SynTeam. The idea of the local search is to perform a small (beneficial) change in the current solution, enabling the algorithm to eventually escape from local optima. In detail, this works as follows. We start from a first team and a first employee in partition P_{best}. We iterate over the teams and the employees within those teams (lines 1 to 4). In every iteration, we swap two currently selected employees and thus, create a new partition P by swapping two employees between teams (line 5). If the synergistic utility contribution of these new teams to the current best partition P_{best} is larger than the utility contribution of the original teams (line 8), we return a newly created partition (line 9). We follow a greedy approach and local search is stopped as soon as a swap yielding a better partition is found. If no better partition is found, we return the current best partition P_{best} (line 15).

5.3 Computational Results

In this section we experimentally evaluate the two algorithms presented in this work: the optimal solver from Section 5.2.3 and SynTeam. The main goals of this empirical evaluation are as follows:

1. to evaluate the algorithms' runtime as team sizes and number of employees increase;

2. to compare the quality of the approximate solutions obtained by SynTeam;

3. to evaluate the anytime performance and guarantees that SynTeam can provide with respect to STCPSolver.

5.3.1 Settings

We used IBM ILOG CPLEX Optimization Studio v12.7.1 [IBM, 2017].

During the experiments, we used the following data and settings for SynTeam:

- **Employees** (Definition 4.2): We generated the data of 102 employees such that each employee has an id, a gender, a personality profile and seven competences with varying competence levels.

- **Task type** (Definition 4.4): The importance given to proficiency was equal to $\lambda = 0.8$. We experimentally assessed this value in the previous Chapter (see Section 4.5). The task type required seven equally important competences (matching with students' competences). Hereby, the required competence levels were all set to $l_i = 1$ with the importance $w_i = 1/7$ for all $c_i \in C$.

- **Task** (Definition 4.6): Team sizes (m) from $[3, 6]$ were considered. We did not perform calculations for higher m as generating the problem form STCPSolver gets more and more costly with increasing m.

- **Evaluating Team Proficiency** (Subsection 4.4.1): By requiring the highest levels of competence, no team member can be over-proficient. Hence, we set the penalty given to the under-proficiency to the maximum value ($\upsilon = 1$). However, the results of this study would not have been affected by any other setting.
- **Evaluating Team Congeniality** (Subsection 4.4.2): We analytically assessed that in order to balance personality requirements (make them equally relevant), the importance given to the different components of the team congeniality function should be as follows: (1) $\alpha = 0.19$, (2) $\beta = 3 \cdot \alpha$, (3) $\gamma = 0.1$.

We chose the following parameters for SynTeam (see algorithm 2):

1. **Maximum number of iterations without any improvement (n_r)**: We wanted to give SynTeam a chance to visit all teams at least once without revisiting the same teams too many times. Hence, we decided to make the value of n_r dependent on the value of b, which is the total number of different teams in a partition. We experimentally assessed how the quality of SynTeam solutions improves over time. In this context, we observed that the total number of teams multiplied by 1.5 — that is, $n_r = 1.5 \cdot b$ — offered a good compromise.
2. **Frequency of the local search (n_l)**: We used the maximum number of iterations without any improvement (n_r) to assess if the algorithm is stuck in a local optimum. We empirically observed that, after performing approximately $\frac{n_r}{6}$ random team re-compositions without any improvement, the probability of finding an improvement was low. Hence, we set the frequency to perform local search to $\frac{n_r}{6}$—as the frequency for performing the local search.

5.3.2 Computational Results

Our computational experiments were ran on the 4 cores of a machine with an Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz and 15 GB of system memory. We measured the runtime as well as the quality obtained over time by both algorithms. STCPSolver used eight threads to perform the calculations, while SynTeam only used one thread. The results are shown in detail in the following subsections.

5.3.2.1 Runtime Analysis

Figure 5.1 shows the performance, in terms of the total running time, of SynTeam and STCPSolver for different teams as the number of employees increases ($m \in \{3, 4, 5, 6\}$). Twenty runs were performed for each configuration, and the average total run times as well as the standard deviations were recorded. Notice that with increasing team size m, generating the problem for STCPSolver becomes

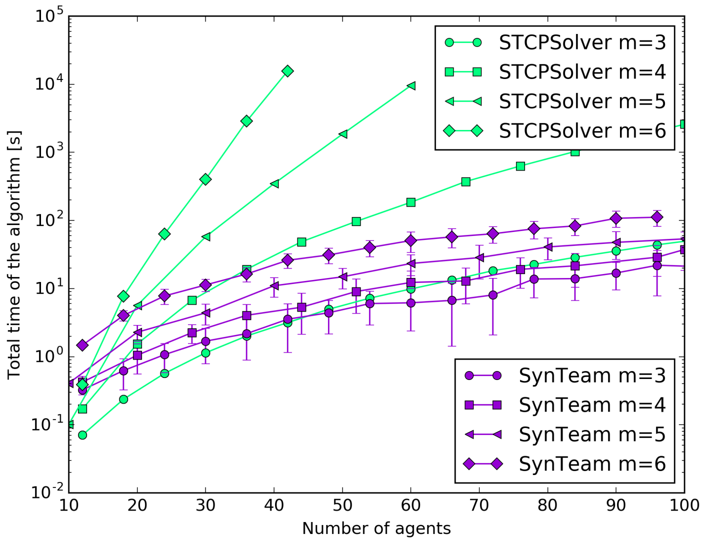


Figure 5.1: The total runtime of the SynTeam and STCPSolver algorithms for different team sizes.

more and more costly (see the last paragraph of subsection 5.2.2 for the equation to calculate the total number of partitions). Therefore, for STCPSolver we were only able to do the calculations for up to 100 employees in the case of $m \in \{3, 4\}$, for up to 60 employees in the case of $m = 5$, and for up to 42 employees in the case of $m = 6$.

We observe that the runtime of STCPSolver increases drastically with an increasing number of employees (n) and coalition size (m). Note that for team size $m = 6$ and $n = 42$ employees, SynTeam ran using one thread is at least 3 orders of magnitude faster than STCPSolver ran using eight threads. This shows the utility of the SynTeam algorithm for team composition problems STCPSolver cannot solve easily.

In order to understand better this result, we performed another experiment. That is, we divided the total runtime of both algorithms into two parts. In case of STCPSolver we wanted to check how long it was taking to: (1a) prepare input for LP Solver and (2a) solve a problem. We compared this runtime results with two parts of the SynTeam algorithm, that is, accordingly: (1b) generating a first team partition, and (2b) improving the team partition.

Figure 5.2(Left) shows the first comparison ((1a) vs. (1b)). Note that the time needed to prepare the input for the STCPSolver is a substantial part of the total time needed to solve a problem. Since SynTeam simply generates a random partition, the time needed to obtain a first solution is minimal. Figure 5.2(Right) shows the second comparison ((2a) vs. (2b)). We observe that — even in this case — SynTeam is more efficient for larger instances (team sizes m

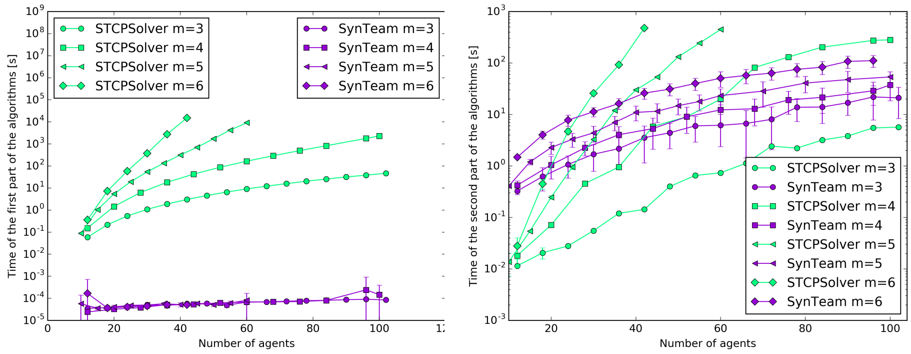


Figure 5.2: (Left) The runtime to obtain first solution for SynTeam (i.e. a random solution) versus file preparation for STCPSolver (i.e. computing the synergistic values and generating a linear programming file). (Right) Runtime of SynTeam versus the time needed to obtain the solution using Cplex (disregarding problem generation time).

larger than 3 and number of employees n larger than 40).

5.3.2.2 Quality Analysis

We further evaluated the performance of SynTeam by assessing the quality of the obtained solutions.

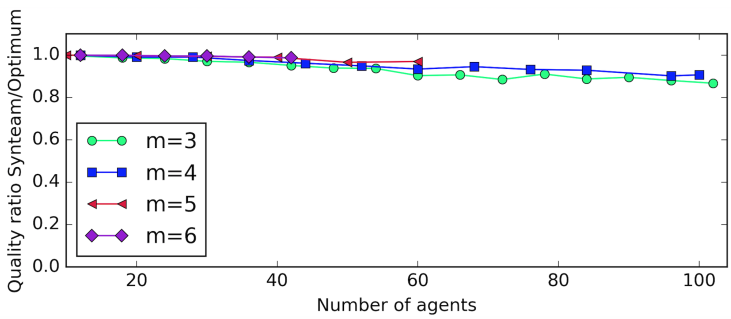


Figure 5.3: The quality ratio of the SynTeam algorithm.

For each case described in subsection 5.3.2.1, we calculated the optimality ratio. Specifically, we divided the solution obtained by SynTeam by the optimal solution calculated by STCPSolver. Figure 5.3 illustrates the quality ratio with respect to the number of employees and team sizes. The results show that the quality of approximate solutions decreases with the number of employees and team sizes but it is always above 87%.

Finally, we aimed at studying the anytime performance of SynTeam and STCPSolver. For this purpose, we chose one the configuration with $n = 45$

employees and a required team size of $m = 5$. We chose this configuration because it is still in the region of problems that STCPSolver could afford. Figure 5.4 shows the evolution of the best found solutions over time (in terms of the quality ratio, that is, divided by the best known solution) for both algorithms. Note that problem generation time required by STCPSolver is not shown, and hence we only plot the time required by cplex.

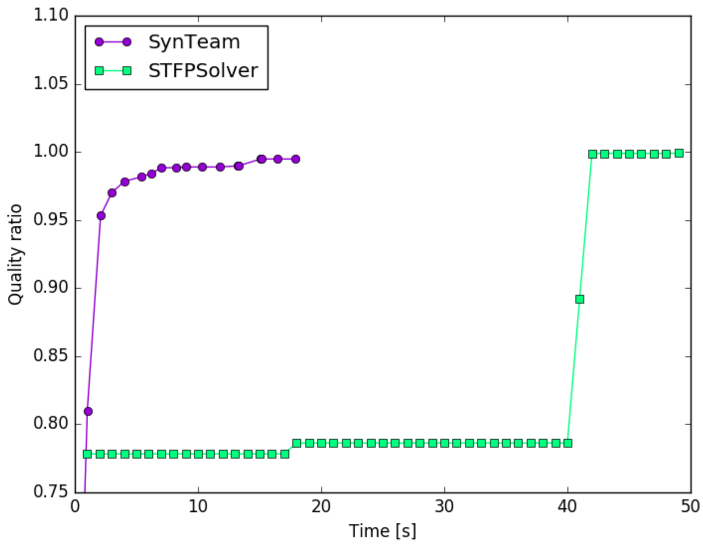


Figure 5.4: Anytime performance (in terms of the quality ratio) of SynTeam and STCPSolver for the case $n = 45$ and $m = 5$.

It can be observed that SynTeam is able to provide very good solutions (quality ratio of over 95%) in less than 3 seconds, while STCPSolver needs approximately 700 seconds (preprocessing time plus solving time) to come up with a first, low-quality solution.

To conclude, in order to reach optimality STCPSolver requires 233 times the time required by SynTeam to obtain solutions very close to optimality.

5.4 Experimental Results

In this section we describe real-life experiments that we performed in several secondary schools in Catalonia that show the effectiveness of our methods.

5.4.1 First Experiment - Final Group Assignment

In this subsection we discuss the details of our first experiment. This experiment took place a year after the experiment presented in Section 4.5 (i.e. in 2017) in the same state school near Barcelona — “Institut Torras i Bages”.

Student and team data: The experiment was performed upon four groups of students (98 students in total): '1r ESO A' (25 students), '1r ESO B' (25 students), '1r ESO C' (22 students) and '1r ESO E' (26 students).

Task type: Similarly to the experiment described in Section 4.5, students were asked to undertake the set of interdisciplinary activities ("Treball de Síntesi"), which is an obligatory exam performed at the end of each year of the secondary education curriculum in Catalonia. In detail, students were asked to create a tourist brochure of their city with all details (collect the information about the city architecture, history, cuisine, main festivals, design the logo, design the brochure, translate parts of the brochure to English). Formally, the task type $\{(c_i, l_i, w_i)\}_{i \in [1,7]}$ had seven equally important competences, $w_i = 1/7$, with a maximally competence level requirement, $l_i = 1$.

Team size: We divided each classroom into teams of size three.

Measuring Personality: Using computers and/or mobile phones, students answered the Post-Jungian Personality questionnaire in Subsection 3.1.2. The full questionnaire is provided in Appendix B.

Competence measure: We measured students' intelligences using a self evaluation test introduced by Rice [Rice, 2013] and translated it into local language (Spanish) consulting school psychologist. The full test and its' translation is provided in Appendix A.

Students knew that the purpose of personality and competence questionnaires was to generate heterogeneous teams, understood the task, and filled in the tests as requested.

Team size: We divided each classroom into teams of size three.

Performance evaluation: Similarly to the procedure from 2016 discussed in Section 4.5.2, students worked in teams and at the end of every activity presented their work in front of a panel of three teachers that assessed the content, presentation and cooperation between team members using a standardized rubric on a scale between 1 and 10.

The procedure:

- We split each class into two halves of similar size using random sampling;
- We partitioned one of the halves into triplets by SynTeam with $\lambda = 0.8$ as learnt in the previous experiment (16 teams in total). The other half was divided by the expert method (15 teams in total);
- All teams performed "Treball de Síntesi" and we collected the final marks of students.

The results: We compared the marks obtained by students. Specifically, we calculated the geometrical average of marks for the teams in each partition. We used geometric average to penalise more the partition that are imbalanced (i.e. the variance in team performance is high). The teams composed by SynTeam obtained 8.1 in the scale between $[1, 10]$, while teams composed by the expert method achieved only 7.3. The relative improvement measured by the difference between two geometric averages and divided by the possible improvement is equal to 29.2%. Hence, we observe that teams composed by SynTeam perform better than the teams composed by the teacher method.

5.4.2 Second Experiment - Scratch programming Task

We performed another experiment to evaluate the effectiveness of our algorithm, when it comes to more technical areas, such as programming.

Place of the experiment: This study took place in three different schools in Catalonia, that is: "Institut Broggi", "Institut Olorda" and "Institut Torras i Bages".

Time of the experiment: This experiment took place between March and November 2017.

Student and team data: The experiment was performed upon five groups of students in ages between 14 and 15 (154 students in total). Specifically, "Institut Broggi" (55 students), "Institut Olorda" (24 students) and "Institut Torras i Bages" (75 students).

Task: The experiment was performed during 2-hour technology classes, where students had to create a game, a story or an animation using the Scratch programming language (https://scratch.mit.edu/).

Personality and Competence test: Similarly to the first experiment, we followed the self-evaluation questionnaires. The personality method is described in Section 3.1 and the full personality questionnaire is provided in Appendix B. The competence method is described in Section 3.2 and the full competence questionnaire is provided in Appendix A.

Team size: We divided each classroom into teams of size two, so that students were able to work in one computer together.

Performance evaluation: An independent Scratch expert that did not know the source of the teamings evaluated the performance of each team following a standardized evaluation form.

The procedure:

- We split each class into two halves of similar size using random sampling;
- We partitioned one of the halves into duets by SynTeam with $\lambda = 0.8$ as learnt in the previous experiment (38 teams in total). The other half was divided by the expert method (37 teams in total);
- All teams performed the task and we collected the final marks of students.

The results: We calculated a geometric average obtained by all teams within each method, which is equal to 5.87 for teams composed using SynTeam and 4.47 for teams composed by the expert method. The relative improvement measured by the difference between two geometric averages and divided by the possible improvement is equal to 25.3%. The observed result is statistically significant (p-value= 0.04). Hence, we observe that again teams composed by SynTeam achieved better performance than the teams composed by teachers.

Finally, table 5.1 shows the summary of both experiments.

Experiment	**Geometric Mean SynTeam Teams**	**Geometric Mean Teacher Teams**	**Relative Improvement**	**#Students**
Final Group Assignment	8.1	7.3	29.2%	98
Scratch Programming	5.87	4.47	25.33%	154

Table 5.1: Relative improvement of SynTeam with respect to teacher method.

5.5 Conclusions

In this chapter we considered the Synergistic Team Composition Problem (STCP) and proposed both an optimal and an approximate solution to this problem. First, we discussed an algorithm to optimally solve the STCP called STCPSolver. When we noticed that the algorithm is only effective for small instances of the problem, we developed SynTeam, a greedy algorithm for partitioning groups of humans into proficient, gender, psychologically and size balanced teams, which yields a good, but not necessarily optimum solution. Our computational evaluation shows that SynTeam outperforms STCPSolver, for the higher number of employees and team sizes. Finally, SynTeam algorithm provides approximate solutions with good quality guarantees (i.e., up to 87%). Next, we performed two different real-life experiments in education scenario with the total of 252 students. The results show teams composed by SynTeam perform better than teams composed by a tutor that knows the students — their background, competences, social and cognitive capabilities. We were not able to benchmark our algorithm against random teamings as composing teams expecting to perform worse raises an ethical question about fairness of students evaluation (final marks might be worse) and may cause drop in students' engagement.

To our knowledge, SynTeam is the first computational model to build synergistic teams based on competences, personality diversity and gender balance.

The algorithm composes teams in a purely automatic way without consulting experts, which is a huge advantage for environments where there is a lack of experts.

Even though the performance evaluation of teams was done by experts who did not know the source of the teaming, the possible biases still exist (like favourite or disliked students). Additionally, an evaluation process costs time. Henceforth, in the next chapter we propose an algorithm to reduce possible biases as well as help assessors in their evaluation load. This is specially critical when assessors face the challenge of evaluating large quantities of individuals as needed in big organisations.

Chapter 6

Individual and Team Assessment

In this Chapter, we present an algorithm for *Performance Assessment* as part of our management organisational workflow presented in Chapter 1 (shown in figure 1.1).

In many areas of our lives we are used to the process of assessing and being assessed. We pass exams at the University, we go through job interviews, we undergo research project reviews, etc. In organisations, from one side, appraisals are the means to an individual's career development by helping identify and set goals for the employee, recognize progress over time, identify problem areas and motivate. Also, accurate appraisal of individuals' attributes (competences, motivation, stress rate, performance etc.) is one clear way to gain knowledge about employees that can be used as an input for team processes. This knowledge can help in composing effective teams, predict a team success for a given task as well as discover the necessity for adding a new employee in a team. Yet, when reviewing the literature on competence and performance appraisals, there appears to be no single best process that is widely used in organisations [Anderson et al., 2001].

When it comes to the Artificial Intelligence Literature, research has focused on the assessment process for long and a number of algorithms have been developed to assist in assessing the performance of humans or artificial agents. Indeed, a large number of trust and reputation models has been proposed [Alfaro and Shavlovsky, 2013; Piech et al., 2013; Walsh, 2014; Wu et al., 2015a; Zhang et al., 2007].

Surprisingly, to our knowledge, no significant effort has been put in the development of algorithms that use *judgment* information over such assessments. We consider exam marks unjust, interview outcomes biased, and review reports unfair, and we normally comment about these opinions on our performance with friends and relatives. We think that this kind of information is very important as it can be key to build the reputation of assessors. A bad assessor can be detected by the assessing community if they were allowed to simply express their opinions about the bad assessor. Actually, in many social networks this kind of information is collected ("was this recommendation useful to you?"), and presented to users. However, how the sites use this information to rank recommendations is never clearly explained if it is used at all.

Similarly, in the area of multiagent systems, agents' performance is key to build teams and coalitions [Osman et al., 2013]. Team formation and coalition formation are key for many applications related to multiagent cooperation, e.g. RoboCup rescue teams [Nair et al., 2003; Ramchurn et al., 2010], Unmanned Aerial Vehicles (UAVs) operations [Haque et al., 2013], or team formation in social networks [Lappas et al., 2009] to name just a few. Both team formation and coalition formation focus on assembling the *best* possible group of agents (be it either a team or a coalition) to accomplish some tasks of interest given some limited resources. Hence, it is key for these algorithms to count on an assessment of the *expected capabilities* of the agents to recruit.

In this chapter we present an algorithm, called *Collaborative judgment* (CJ), which wants to go a step further in the use of peer opinions. CJ takes into account judgments on opinions to build reputation values on assessors and then use them to aggregate the opinions of a group of assessors. In current recommender systems the opinions about an object are often aggregated using weights. When no weights are used, the final opinion is usually an average of all the opinions provided (e.g. Amazon or TripAdvisor). When weights are used the aggregated opinion is a weighted average using self-assigned weights. This is very common in Conference Management Systems like Confmaster or Easychair. Later on, we will compare CJ with the standard algorithm that weighs opinions with the assessors' self-assessments. We will call this simple algorithm *Self-Assessment Weighted Algorithm* (SAWA).

This chapter makes the following contributions. First, we define our ranking algorithm based on collective assessments that uses both peer opinions of employees as well as peer judgments over these opinions. Next, we apply it to the case of scientific paper assessment. We compare paper evaluations' accuracy with the currently most used paper evaluation method: the average of opinions weighted by peer self-confidence. Finally, we experimentally compare the *partial ranking* among alternatives produced by both methods and the "actual" ranking. The results show that the rankings produced by our algorithm improve those produced with current ranking methods.

The remaining of this chapter is organised as follows. In Section 6.1 we present the ranking algorithm that we benchmark in Section 6.4 against SAWA, presented in Section 6.2. Then, in Section 6.5 we discuss the results and summarise our main achievement and outline our future work.

6.1 CJ: Collaborative judgments Model

The collaborative judgments model (CJ) aggregates peer assessments by weighing each assessment with respect to its reliability, where reliability in this model is referred to as the peer's reputation, and it is based on the judgments that this peer has received.

In this section we detail our CJ algorithm, but first, we introduce the notation, which we will use in the rest of this section.

Definition 6.1. An *Appraisal* is a tuple $\langle P, R, E, o, v \rangle$, where

- $P = \{p_i\}_{i \in \mathcal{P}}$ is a set of objects to be evaluated.
- $R = \{r_j\}_{j \in \mathcal{R}}$ is a set of peers (employees reviewing objects).
- $E = \{e_i\}_{i \in \mathcal{E}} \cup \{\perp\}$ is a totally ordered evaluation space, where $e_i \in \mathbb{N}$ and $e_i < e_j$ iff $i < j$ and $\perp$ stands for the absence of evaluation.
- $o : R \times P \to E$ is a function giving the opinions of peers on objects.
- $v : R \times R \times P \to E$ is a function giving the judgments of peers over opinions on objects. Therefore, a judgment is a peer's opinion about another peer's opinion.

In general we might have different dimensions of evaluation, that is a number of E spaces over which to express opinions and judgments. For instance, originality, soundness, etc. Nonetheless, here for simplicity reasons we will assume that the evaluation of an object is made over a single dimension. Actually, the 'overall' opinion is what is aggregated in real systems.

6.1.1 Collaborative Judgment Algorithm.

The steps of the CJ algorithm applied over an appraisal $\langle P, R, E, o, v \rangle$ are as follows:

Step 1. Compute the *agreement level* between each pair of peers r_i and r_j as a function $a : R \times R \to [0, 1] \cup \{\perp\}$. This computation involves the set of objects jointly reviewed by peers r_i and r_j, which we will formally define as $P_{ij} = \{p_k \in P | o(r_i, p_k) \neq \perp, o(r_j, p_k) \neq \perp\}$. If two peers jointly reviewed objects then the algorithm uses the judgments on the opinions, in case they exist. Otherwise the similarities between the opinions over the common papers are combined. Formally, we compute the agreement level as:

$$a(r_i, r_j) = \begin{cases} \frac{\sum_{p_k \in P_{ij}} s(r_i, r_j, p_k)}{|P_{ij}| \cdot d} & \text{if } P_{ij} \neq \emptyset \\ \perp & \text{otherwise} \end{cases} \tag{6.1}$$

where d is the maximum distance in the evaluation space and:

$$s(r_i, r_j, p_k) = \begin{cases} v(r_i, r_j, p_k) & \text{if } P_{ij} \neq \emptyset \text{ and } v(r_i, r_j, p_k) \neq \perp \\ Sim(o(r_i, p_k), o(r_j, p_k)) & \text{if } P_{ij} \neq \emptyset \text{ and } v(r_i, r_j, p_k) = \perp \\ \perp & \text{otherwise} \end{cases} \tag{6.2}$$

Sim stands for an appropriate similarity measure. When no explicit judgments are given, we use the similarity between opinions as a heuristic of their values. This is based on the following assumption: the more similar a review is to my opinion, the better I am bound to judge that opinion.

Step 2. Compute a complete *Trust Graph* as an adjacency function matrix $C = \{c_{ij}\}_{i,j \in R}$ such that:

$$c(r_i, r_j) = \begin{cases} a(r_i, r_j) & \text{if } a(r_i, r_j) \neq \perp \\ \max\limits_{h \in chains(r_i, r_j)} \prod\limits_{(k,k') \in h} a(r_k, r_{k'}) & \text{otherwise} \end{cases} \tag{6.3}$$

where $chains(r_i, r_j)$ is the set of sequences of peer indexes connecting i and j. Formally, a chain h between peers i and j is a sequence $\langle l_1, \ldots, l_{n_h} \rangle$ such that $l_1 = i$, $l_{n_h} = j$, and $a(r_k, r_{k+1}) \neq \perp$ for each pair $(k, k+1)$ of consecutive values in the sequence. To compute this step we use a version of Dijkstra's algorithm that instead of looking for the shortest path (using $+$ and min as mathematical operations), it looks for the path with the largest edge product (using $\cdot$ and max as mathematical operators). The running time of the Dijkstra algorithm can take $O(n \log n)$, where $n = |R|$, if using priority queues [Cormen et al., 2001].

Step 3. Compute a *reputation* for each peer in R, $\{t_i\}_{i \in R}$. With this aim, we follow the notion of transitive trust: If a peer i trusts any peer j, it would also trust the peers trusted by j. Since this principle is employed by the Eigentrust algorithm [Kamvar et al., 2003], we use it to compute peer reputations. The use of Eigentrust allows us to obtain a global trust value for each peer by the repeated and iterative multiplication and aggregation of reputation values until the trust grades for all employees converge to stable values. Note that the trust graph generated in step 2 is aperiodic and strongly connected as required by the Eigentrust algorithm. Furthermore, we normalise the powers of the matrix C at each step to ensure its convergence. In vectorial notation, the trust vector is assessed as $\bar{t} = \lim_{k \to \infty} \bar{t}^{k+1}$ with $\bar{t}^{k+1} = C^T \bar{t}^k$ and $\bar{t}^0 = \bar{e}$ being $\bar{e}_i = 1/|\bar{e}|$. The complexity of the Eigentrust algorithm used in this step is $O(|R|^2)$.

Step 4. Compute the *collective opinion* on each object as a weighted average of the opinions of those that expressed an opinion on the object. In other words, given an object p_j, we only consider the opinions of those peers that reviewed p_j, which we formally define as $R_j \subseteq R, R_j = \{r \in R | o(r, p_j) \neq \perp\}$. We can then compute the collective opinion on an object p_j as a weighted average of the opinions of the peers in R_j using as weights the peers' reputations. Finally, the collective opinion computed by our collaborative judgment algorithm for an object p_j, noted as $o_{CJ}(p_j)$, is:

$$o_{CJ}(p_j) = \frac{\sum_{r \in R_j} \bar{t}_r \cdot o(r, p_j)}{\sum_{r \in R_j} \bar{t}_r} \tag{6.4}$$

where $\bar{t}_r$ stands for the reputation value of peer r.

Step 5. Generate a partial ranking based on the set of collective opinions $O_{CJ}(P)$. CJ sorts objects in descending order by the collective opinion values. Thus, the object with the highest value of collective opinion gets the first ranking position. Objects with equal collective opinion receive the same ranking number, and the object(s) on the next position receive the immediately following ranking number (i.e. bucket index). The procedure continues until CJ assigns bucket indexes to all objects.

6.2 The Self-Assessment Weighted Algorithm

A conference management system is a web-based application that supports, *inter alia*, the evaluation and selection of articles for scientific purposes (mainly conferences and to some degree journals). The most common approach to paper evaluation used in systems such as Confmaster or Easychair is as follows:

1. Assign every article to (normally) three peers based on either keywords distinguished by using word frequency analysis, and eventually their preferences expressed as bids.

2. Ask each peer to assess (give an opinion on) each of their assigned papers and also assess their own confidence on each evaluation.

3. Determine the overall opinion on each paper as a weighted average of the opinions of the peers of the paper considering their *self-assessed confidences* as weights.

4. Build a (partial) ranking of articles based on the overall opinions.

We will refer to the algorithm above as the *self-Assessment Weighted Algorithm* (SAWA). Next, we formalise how step 3 in SAWA computes the overall opinion on each paper. We assume that a function $\kappa : R \times P \mapsto [0, 1]$ keeps how confident each peer feels about their opinion on a paper. Then, given a paper p_j evaluated by a set of peers R_j, SAWA computes the aggregated opinion on the paper as:

$$o_{SAWA}(p_j) = \frac{\sum_{r \in R_j} \kappa(r, p_j) \cdot o(r, p_j)}{\sum_{r \in R_j} \kappa(r, p_j)} \tag{6.5}$$

We rank articles in descending order according to values $O_{SAWA}(P)$ in the same way as in CJ.

6.3 Motivating Example.

The purpose of this subsection is to illustrate how the CJ and SAWA algorithms described in subsections 6.1 and 6.2 work to produce paper rankings. Before that, we introduce some matrix notation that will help us describe CJ's operation in a concise manner. Thus, let $O : |P| \times |R|$ be the opinion matrix; $A : |R| \times |R|$ be the agreement level matrix; $V_k : |R| \times |R|$ the individual judgment matrix for paper p_k containing only direct judgments of peers; $S_k : |R| \times |R|$ the judgment matrix for paper p_k; $C : |R| \times |R|$ the trust matrix; and $\bar{t}$ the reputation vector.

Now, say that there are only four papers to be reviewed $P = \{p_0, p_1, p_2, p_3\}$ and four peers $R = \{r_0, r_1, r_2, r_3\}$ available to give their opinions on papers. Our objective is to choose two top-rated articles out of P and compare CJ and SAWA rankings of the papers in P. We consider that peers r_0 and r_1 are qualified, which means that they can recognize the value of a paper and rate it adequately. peers r_2 and r_3 provide unfair opinions as they are incompetent, but they can distinguish between a good and a bad review, namely they can judge correctly. Every article is assigned to three peers as follows:

- p_0 is assigned to $\{r_0, r_1, r_2\}$,
- p_1 is assigned to $\{r_1, r_2, r_3\}$,
- p_2 is assigned to $\{r_1, r_2, r_3\}$,
- p_3 is assigned to $\{r_0, r_1, r_3\}$

We assume that all peers but r_3 complete their reviews. Peer r_3 did not evaluate article p_3. Based on the collected reviews, the opinion matrix O looks as follows:

$$O = \begin{bmatrix} 0.1 & 0.1 & 0.7 & \perp \\ \perp & 0.5 & 0.5 & 0.9 \\ \perp & 0.6 & 0.7 & 0.4 \\ 0.9 & 0.9 & \perp & \perp \end{bmatrix}$$

For instance, the opinion of peer r_3 on paper p_2 is 0.4, namely the value of $O[2, 3]$.

Besides reviews, each peer evaluates their own confidence on each of their reviews and judges the reviews of other peers whenever they have papers in common. In other words, given a paper p_k, the peers in R_k judge one another. Thus, each individual judgment matrix V_k will contain each peer self-assessment together with the peers' judgments on other reviews of p_k. Say that the individual judgment matrices in our example are defined as follows:

$$V_0 = \begin{bmatrix} 0.9 & 1.0 & 0.8 & \perp \\ 1.0 & 1.0 & 0.8 & \perp \\ 0.2 & 0.2 & 0.7 & \perp \\ \perp & \perp & \perp & \perp \end{bmatrix} V_1 = \begin{bmatrix} \perp & \perp & \perp & \perp \\ \perp & 0.9 & 1.0 & 0.6 \\ \perp & 0.3 & 1.0 & 0.7 \\ \perp & 0.2 & 0.7 & 0.8 \end{bmatrix}$$

$$V_2 = \begin{bmatrix} \perp & \perp & \perp & \perp \\ \perp & 0.9 & 1.0 & 0.7 \\ \perp & 0.5 & 1.0 & 0.8 \\ \perp & 0.3 & 0.1 & 0.6 \end{bmatrix} V_3 = \begin{bmatrix} 1.0 & 0.9 & \perp & \perp \\ \perp & 1.0 & \perp & \perp \\ \perp & \perp & \perp & \perp \\ \perp & \perp & \perp & \perp \end{bmatrix}$$

For instance, consider the individual judgment matrix V_0. Peer r_2 indicates that their self-assessed confidence on their review of paper p_0 is 0.7, namely the value of $V_0[2,2]$. Furthermore, peer r_2 judges the review made by r_1 with a 0.2 value, namely the value of $V_0[2,1]$.

At this point, we count on all the input information required by CJ and SAWA to perform paper assessment and produce paper rankings. Next, in sections 6.3.1 and 6.3.2 we illustrate CJ's and SAWA's operations respectively, while in section 6.3.3 we compare the rankings produced by both algorithms.

6.3.1 The collaborative judgment algorithm at work

We follow the steps for the CJ algorithm described in section 6.1:

Step 1. Compute the *agreement level* between peers. This requires that we compute first the judgment matrices S_0, S_1, S_2, and S_3 for the papers in P using equation 6.2. CJ sets $S_0 = V_0$, $S_1 = V_1$, and $S_2 = V_2$. As to V_3, it finds that there is a missing judgment of r_1 about r_0 when it comes to opinion about paper p_3. Then, it calculates this missing judgment by considering the difference in opinions between r_0 and r_1 on paper p_3 using the following similarity measure:

$$Sim(o(r_i, p_k), o(r_j, p_k)) = 1 - |o(r_i, p_k) - o(r_j, p_k)|$$

Hence, the final matrix of judgments for paper p_3 looks as follows:

$$S_3 = \begin{bmatrix} 1.0 & 0.9 & \perp & \perp \\ 1.0 & 1.0 & \perp & \perp \\ \perp & \perp & \perp & \perp \\ \perp & \perp & \perp & \perp \end{bmatrix}$$

Now CJ can employ equation 6.1 to calculate the agreement level matrix:

$$A = \begin{bmatrix} 0.95 & 0.95 & 0.80 & \perp \\ 1.00 & 0.95 & 0.93 & 0.65 \\ 0.20 & 0.33 & 0.90 & 0.75 \\ \perp & 0.25 & 0.40 & 0.70 \end{bmatrix}$$

Step 2. Compute a complete *trust graph*. If we look at equation 6.3, we observe that we can readily obtain most trust values from the agreement matrix A. In fact, we only miss the trust values between r_0 and r_3 (notice that

$a[0,3] = \perp$ and $a[3,0] = \perp$ in the agreement matrix A). Therefore, CJ only has to compute $c(r_0, r_3)$ and $c(r_3, r_0)$. Recall from section 6.1 that the missing trust value for a pair or peers r_i and r_j is computed by finding the chain (path) of peers connecting r_i and r_j with maximum trust product. Figure 6.1 shows a graph-based representation of the agreement level matrix A, nodes stand for peers and a directed edge from r_i to r_j is labeled with the agreement level between r_i on r_j, namely $a(r_i, r_j)$.

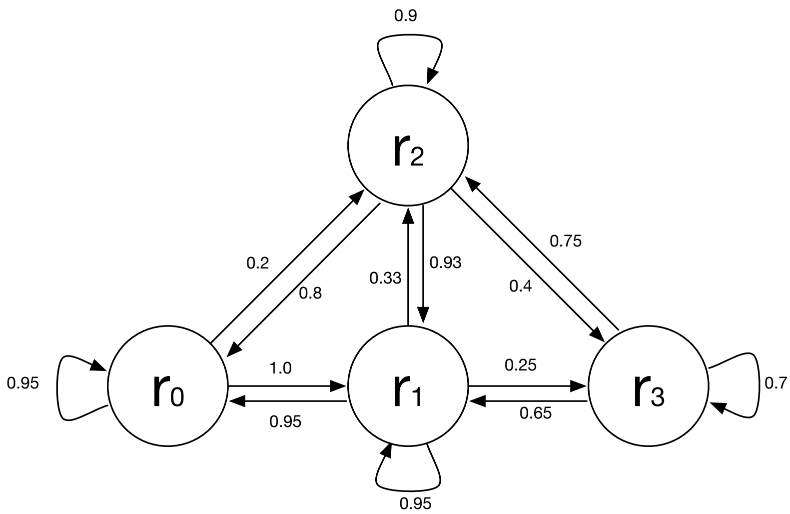


Figure 6.1: The graph representing the agreement level between peers.

Our algorithm finds that the missing agreement levels between r_3 and r_0 are:

- $c(r_0, r_3) = 1.0 \cdot 0.25 = 0.25$ because the chain with maximum trust product is $\langle r_0, r_1, r_3 \rangle$
- $c(r_3, r_0) = 0.75 \cdot 0.93 \cdot 0.95 = 0.663$ because the chain with maximum trust product is $\langle r_3, r_2, r_1, r_0 \rangle$.

By putting together the values of the agreement level matrix and the missing agreement levels $c(r_0, r_3)$ and $c(r_3, r_0)$, we finally obtain the trust matrix C:

$$C = \begin{bmatrix} 0.95 & 0.95 & 0.80 & 0.66 \\ 1.00 & 0.95 & 0.93 & 0.65 \\ 0.20 & 0.33 & 0.90 & 0.75 \\ 0.25 & 0.25 & 0.40 & 0.70 \end{bmatrix}$$

Step 3. Compute a *reputation* value for each peer in R by using Eigentrust. Finally, the algorithm computes the reputation values of each employee by applying Eigentrust with the C matrix obtained at step 2 as an input.[1] Eigentrust converges to the following reputation vector:

$$\bar{t} = \begin{bmatrix}0.344 & 0.358 & 0.169 & 0.129\end{bmatrix}^T$$

Each row in $\bar{t}$ represents a reputation value for each one of the peers in R. We observe that the reputations of r_2 and r_3 are 0.169 and 0.129 respectively. Therefore, CJ found that these two peers are not competent.

Step 4. Compute the *collective opinion* on objects as a weighted average of the opinions of those that expressed an opinion. Having assessed the employees' reputation, CJ can calculate the collective opinion for each paper using equation 6.4. The resulting opinions for each paper are shown in figure 6.2.

Step 5. Generate a partial ranking based on the set of collective opinions $O_{CJ}(P)$. CJ generates a paper ranking that comes from ordering papers according to the opinion values in descending order, as shown in figure 6.3.

6.3.2 SAWA at work

SAWA computes the opinion on each paper by combining the values in the opinion matrix O with the self-assessed confidence of each peer in the individual judgment matrices V_0, V_1, V_2, and V_3, namely with the value in the diagonals of these matrices. For instance, let us calculate the opinion on article p_2. This requires the opinions of peers r_1, r_2, and r_3 on the article ($O[2,1] = 0.6$, $O[2,2] = 0.7$, and $O[2,3] = 0.4$). It also requires the confidence values of those peers in their reviews, which are contained in matrix V_2: $V_2[1,1] = 0.9$, $V_2[2,2] = 1.0$, and $V_2[3,3] = 0.6$ are the confidence values of r_1, r_2, and r_3 respectively on their own reviews on p_2. Now, using equation 6.5, SAWA assesses the opinion on p_2 as a weighted average of opinions using confidence values as follows:

$$o_{SAWA}(p_2) = \frac{0.6 \cdot 0.9 + 0.7 \cdot 1.0 + 0.4 \cdot 0.6}{0.9 + 1.0 + 0.6} = 0.592$$

The opinions for the rest of articles are shown in Figure 6.2 below.

6.3.3 Comparing rankings

Next we compare the paper rankings produced by CJ and SAWA with the ranking resulting from an "oracle" that knows the true quality of the papers. Figure 6.3 shows the produced rankings based on the opinions in Figure 6.2.

[1] The matrix gets transposed to be used by the Eigentrust algorithm, therefore each peer reputation is represented by one column

Figure 6.2: Opinions obtained by CJ and SAWA per paper together with their true quality values.

	p0	p1	p2	p3
CJ	0.217	0.579	0.586	0.900
SAWA	0.262	0.619	0.592	0.900
True Quality	0.100	0.500	0.600	0.900

Figure 6.3: Ranking produced by CJ and SAWA along with the ranking resulting from the papers' true qualities.

	Ranking
CJ	{p3}, {p2}, {p1}, {p0}
SAWA	{p3}, {p1}, {p2}, {p0}
True Ranking	{p3}, {p2}, {p1},{p0}

We observe that the ranking produced by CJ is the same as the oracle's, while SAWA yields a different ranking. This is because CJ exploited judgment information to find out that peers r_2 and r_3 are incompetent (their reputation values are the lowest ones in $\bar{t}$). This reduced the significance of their opinions when evaluating article p_1, and also increased the importance of the opinion of peer r_1, who is a good reviewer. As a result, the opinion on p_2 is larger than the opinion on p_1. Contrarily, SAWA valued article p_1 better than p_2. This is because peer r_3, a bad reviewer, evaluated better p_1 than p_2 and reported a high self-assessed confidence value. As a result, p_1's opinion outperformed p_2's despite p_2 true quality is higher.

Our example tells us that a more informed algorithm (by adding judgments of opinions) helped us discriminate good assessments from bad assessments. By all means this is just a toy example intended to illustrate our algorithm. In what follows we perform a more substantial evaluation.

Before that, notice that our example only considered full rankings instead of partial rankings (rankings with ties) to ease comprehension.

6.4 Experimental Evaluation

The purpose of this section is to evaluate the CJ algorithm via simulation. Here, we will particularize the problem of peer judgment to the case of Conference Paper reviewing. However, the algorithm is general, and can be of use when evaluating the competences of employees as well as the individual and team performance.

With this aim, we benchmark CJ and SAWA against an "oracle" that knows the true quality of papers. Our analysis will measure:

- the accuracy of the opinions and rankings produced by CJ and SAWA;
- the robustness of CJ against bad reviewers; and
- the sensitivity of global trust to bad reviewers.

Our study will confirm that CJ is the algorithm of choice to compute rankings on objects taking into account peer opinions. Next, in section 6.4 we formulate the hypothesis that our experiments pursue to validate. Section 6.4.1 describes our experimental settings and section 6.4.2 dissects the results of the three experiments providing support to our hypotheses.

Hypotheses In order to demonstrate that CJ is the algorithm of choice to compute rankings taking into account peer opinions, the experiments that follow focus on validating the next hypotheses:

H1 **CJ evaluations get closer than SAWA's to the true quality of a paper as the number of good reviewers increase.**[2]

H2 **The rankings produced by CJ get closer to the true ranking than SAWA's as the number of good reviewers increase.**

H3 **Ceteris paribus, the better the reviewers, the better the accuracy (in terms of opinions and rankings) of CJ with respect to SAWA.**

6.4.1 Experimental settings

We assume a set $P = \{p_1, \ldots, p_n\}$ of papers and a function for their true quality in a range $[0, 1]$,[3] $q : P \to [0, 1]$. We use the following evaluation space $E = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, which is rather common in the context of paper reviewing.

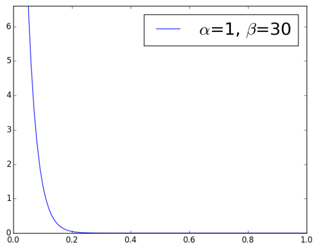
We use beta distributions to model reviewers' opinions and judgments as it is an appropriate distribution to simulate a behaviour that is subject to random variation and is limited on both extremes, i.e. represents processes with natural lower and upper boundaries [Hill and Lewicki, 2006]. Depending on the α and β parameters, the shape of the beta distribution changes substantially (see figure 6.4 below with different configurations of both variables).

We model two types of reviewers: good and bad, with the following behaviour:

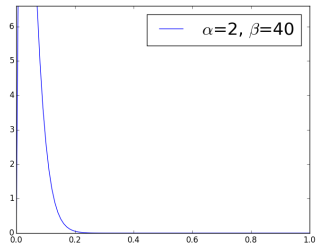
- *Good reviewer.* She provides fair opinions and fair judgments. Her opinion on any paper p_k is always close to its true quality $q(p_k)$. We assume that the absolute value of the difference between the opinion of a reviewer and the true quality of a paper (as a percent) follows a beta distribution,

[2]See next subsection for our representation of a good reviewer.

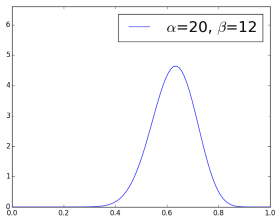
[3]Assessing the true quality of an object may be difficult and it is certainly a domain dependent issue.



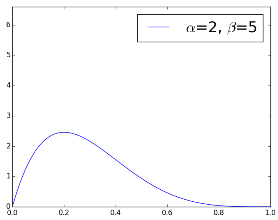
(a) Beta distribution used to model the difference between the opinion of a good reviewer and the true quality



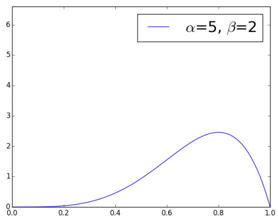
(b) Beta distribution used to model a good reviewer judging a bad reviewer



(c) Beta distribution used to model a bad reviewer opinion



(d) Beta distribution used to model a bad reviewer judging a good reviewer



(e) Beta distribution used to model a bad reviewer judging a bad reviewer

Figure 6.4: Beta Distribution for different configurations of α and β parameters

$Beta(\alpha, \beta)$, very positively skewed, for instance with $\alpha = 1$ and $\beta = 30$. For each paper p_k reviewed by a good reviewer, we sample the reviewer's associated beta distribution for a percentage difference, apply it to the paper quality $q(p_k)$ (up or down randomly) and round the result to fit an element in E. Her judgments on someone's opinion are close to 0 if that opinion is far from the true quality of the paper, and close to 1 otherwise. We implement this as the following function:

$$v(r_i, r_j, p_k) = 1 - |o(r_j, p_k) - q(p_k)|$$

and self-judgments from $Beta(5, 2)$, slightly negatively skewed.

We assume that when a good reviewer judges a bad reviewer she samples a value in E from a beta distribution rather positively skewed: $Beta(2, 40)$. The intuition is that good reviewers poorly mark bad reviews.

- *Bad reviewer.* She provides unfair opinions, because she is incompetent, but provides reasonable judgments as she can interpret the opinions of others as being informative or not. Thus, we sample opinions from $Beta(20, 12)$ —rather central with a slight negative skew, judgments for good reviews and self-judgments from $Beta(5,2)$ as for good reviewers —negatively skewed, and judgments on bad reviews from $Beta(2,5)$ —slightly positively skewed. The overall idea is that bad reviewers stay mostly in the central area of the evaluation space.

We use $Sim(x,y) = (|E| - 1 - |\tau(x) - \tau(y)|)/(|E| - 1)$ as a simple linear similarity function where τ is a function that gives the position of an element in the ordered set E.

6.4.2 Results

In this section we present our experimental results using the settings described above.

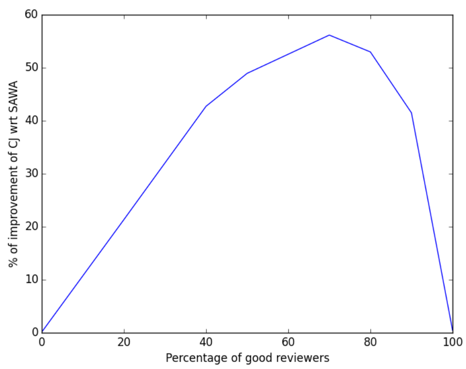


Figure 6.5: Accuracy of opinions: percentage of error improvement of CJ over SAWA.

Analysing the accuracy of opinions Here we consider the accuracy of a collective opinion on a paper as the difference between that opinion and the true quality of the paper. Then we compare the accuracy of the opinions computed by CJ and SAWA as the percentage of good reviewers increases. We compute the accuracy of both CJ and SAWA as the mean absolute error of their opinions with respect to the true qualities using the following expressions:

$$MAE_{CJ} = \frac{\sum_{p \in P} |o_{CJ}(p) - q(p)|}{|P|} \qquad MAE_{SAWA} = \frac{\sum_{p \in P} |o_{SAWA}(p) - q(p)|}{|P|}$$

^{o} where q is a function that yields the true quality of each paper. Figure 6.5 plots the percentage error reduction of CJ with respect to SAWA (computed as $(1 - \frac{MAE_{CJ}}{MAE_{SAWA}}) \cdot 100$) by aggregating the values obtained from 30 runs of each algorithm (each run samples all the distributions and thus generates different collective assessments). Note that CJ outperforms SAWA, as it is much more resilient to bad reviewers. As a matter of fact, as opposed to SAWA that treats all reviewers equally, CJ is designed to detect bad reviewers and diminish the importance of their opinions by the usage of the reputation measure. We observe that CJ's gains become larger than 20% and statistically significant for percentages of good reviewers between 20% and 80% [4]. Therefore, these results support hypotheses H1.

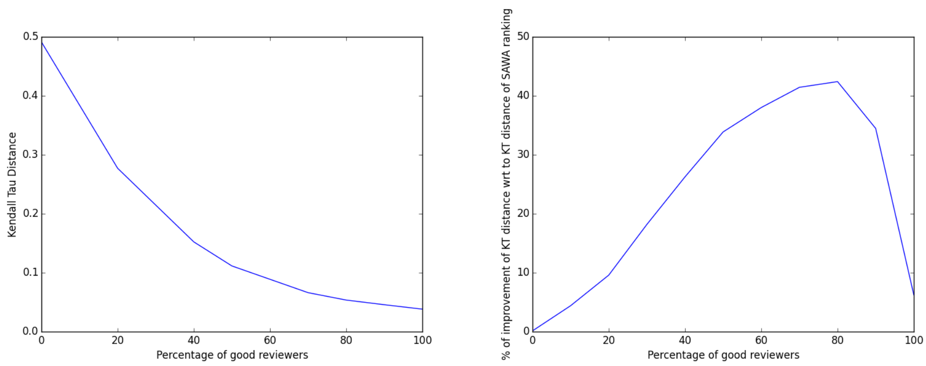


Figure 6.6: (Left) Normalised Kendall Ranking distance calculated for CJ ranking and true ranking of the papers. (Right) Percentage of error decrease measured as a Kendall distance between rankings produced by CJ and SAWA and true ranking of papers for increasing percentages of good reviewers.

Analysing the accuracy of rankings Now we compare the accuracy of the rankings produced by CJ and SAWA with respect to the ranking resulting from the true quality of papers. In order to compare two partial rankings we employ the normalised Kendall distance (see Section 2.1, definition 2.5) with penalty factor $p = 0.5$. We employed the partial rankings resulting from 30 runs of CJ and SAWA. We note by $\sigma_1^{CJ}, \ldots, \sigma_{30}^{CJ}$ the partial rankings produced by CJ by $\sigma_1^{SAWA}, \ldots, \sigma_{30}^{SAWA}$ the partial rankings produced by SAWA, and by σ^q the true ranking. Then, for each partial ranking computed by CJ and SAWA, we compute its normalised Kendall distance with respect to the true ranking. On the one hand, we assess the average Kendall distance of the rankings produced by CJ as $K_{CJ} = \frac{\sum_{i=1}^{30} \tilde{K}^{(0.5)}(\sigma_i^{CJ}, \sigma^q)}{30}$. On the other hand, we assess the average Kendall distance of the rankings produced by SAWA as $K_{SAWA} = \frac{\sum_{i=1}^{30} \tilde{K}^{(0.5)}(\sigma_i^{SAWA}, \sigma^q)}{30}$.

Figure 6.6 (left) plots the average Kendall distance of the rankings produced by CJ with respect to the true ranking, namely K_{CJ}, as the number of good reviewers increases. Note that the lower the distancem, the closer the ranking to

[4]Notice that in systems with no experts, it is nearly impossible to judge a true value of a paper. Similarly, in systems with a very high percentage of good reviewers, the aggregation method is not that important as everyone provides fair opinions. Therefore, it is not crucial for the system to be resilient against bad reviewers.

the true quality ranking. We observe that the distance between CJ rankings and the true ranking quickly decreases as the number of good reviewers increases. Notice that beyond 50% of good reviewers the distance drops below 0.1. That means that CJ can produce rather accurate rankings despite the presence of a large ratio of bad reviewers.

Figure 6.6 (right) shows the accuracy gain of CJ with respect to SAWA. We calculate such accuracy gain as $\frac{K_{SAWA}-K_{CJ}}{K_{SAWA}} \cdot 100$. We observe that the accuracy gain yield by CJ as the number of good reviewers grows, going beyond a 40% gain with 80% good reviewers. Similarly to experiment 6.4.2, the graph clearly shows that CJ performs significantly better even when the number of bad reviewers is high. We see that CJ has been able to discriminate poor assessments, while SAWA treats all reviews equally. We observe also that CJ benefits larger from good reviewers than SAWA.

The results depicted in Figure 6.6 support hypothesis H2.

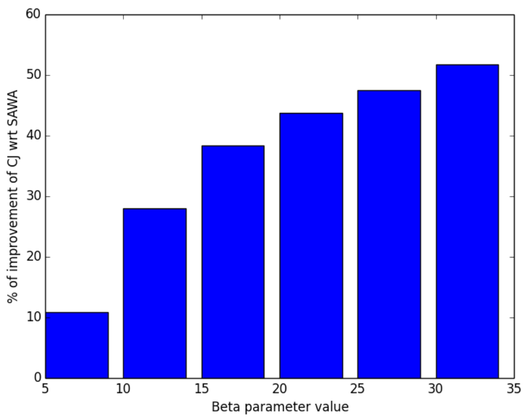


Figure 6.7: Improvement of CJ over SAWA as the reviewers' quality increases (with fixed $\alpha = 1$ and increasing β values). This plot is for a population with 50% good reviewers and 50% bad reviewers.

Analysing the robustness against bad reviewers As mentioned before, we model good reviewers' opinions with a $Beta(\alpha, \beta)$ very positively skewed from which we sample the difference between the reviewer's opinion and the true quality. With $\alpha = 1$ and $\beta > 30$ the expert is frequently telling the true quality in her opinions (specially because we discretise the sampled values into our evaluation space, i.e. almost all the distribution mass is rounded to a distance of 0 with respect to the true quality). In figure 6.7 we plot the improvement of CJ with respect to SAWA for $\alpha = 1$ and increasing values of β (better reviewer behaviour). We observe that the algorithm outperforms SAWA by 10% when reviewer is frequently mistaken ($\beta = 5$). This shows that even when good reviewers give frequently inaccurate opinions, CJ is still able to capture them and increases the importance of their assessments. The improvement asymptotically grows to 51% with increasing quality of the reviewer behaviour. These results support Hypothesis H3.

6.5 Conclusions and Discussion

In this chapter we introduced the algorithm called Collaborative Judgment (CJ). It is a new ranking algorithm that takes into account *peer opinions* of employees as well as *peer judgments* over those opinions. We applied CJ to the use case of scientific paper assessment and we validated it over simulated data. The experiment went as follows. First, we assumed that each paper has a "true quality" value that informs us on how good it is. Second, we modelled two kinds of reviewers such that: (1) good reviewers provide fair opinions and fair judgments, (2) bad reviewers when giving opinions stay mostly in the central area of the evaluation space as they cannot assess true paper quality (since they are incompetent), but provide reasonable judgments as they can evaluate opinions of others as being informative or not. Finally, we compared CJ with the Self-Assessment Weighted Algorithm (SAWA), which is the standard algorithm used in Conference Management Systems. In contrast to SAWA that treats all reviewers equally, CJ is designed to detect biased reviewers and diminish the importance of their opinions using the reputation measure. We were interested in analysing the accuracy of opinions and generated rankings based on those opinions as well as in assessing the robustness against bad reviewers. When it comes to accuracy, we observed that CJ's gains become larger than 20% and statistically significant for percentages of good reviewers between 20% and 80%. Notice that the gains go above 50%, when we have a moderate number of good reviewers (50% − 80%). Similarly, the distance between CJ rankings and the true rankings dropped below 0.1 when we had beyond 50% of good reviewers. We observe that rankings generated by CJ with respect to SAWA are closer to the true ranking and the improvement asymptotically grows to 42% for 80% of good reviewers. These results show that rankings produced by this new algorithm (under (reasonable) assumptions on reviewer behaviour) improve current scientific paper ranking practice. When it comes to robustness, we observe that the algorithm outperforms SAWA by 10% when good reviewers are frequently mistaken. When improving the quality of the good reviewer behaviour, the improvements grow to 51%. These results show that CJ outperforms SAWA, as it is much more resilient to biased reviewers.

The use of this algorithm in the context of employee team composition is key as it provides a sound method to assess the *capabilities* of employees and to measure team performance by observing peer opinions and judgments made by peers.

One issue worth discussing is the feasibility of getting real data to model $q(\cdot)$. We mentioned before that this is obviously a domain dependent issue and that it can be difficult to obtain. In the case of paper review, what is the true quality of a paper? It seems impossible to answer this question. We could get data on impact of papers and assume that impact relates to quality. This can be done for the papers that were accepted and published, but not for those that were rejected. Therefore, the validation of the algorithm results will necessarily be partial. This will always be controversial as the use of any quality metric would always be debatable. It is in this context that our algorithm contributes since the key assumption of our algorithm is: *when there is no clear-cut method to determine the quality of an object, then the true quality can be determined by the social acceptance of the opinions expressed by experts.* Hence, the use of the *best* experts' ranking can be understood as the ranking of the socially most *reputed* experts. Precisely what CJ aims at modelling.

In terms of scalability, the current version of CJ uses Dijstra's algorithm and matrix operations that scale up reasonably well (quadratically), but there are improvements that can be done by distributing the computation as in some versions of Eigentrust.

Another issue worth mentioning is that reviewer quality depends on the particular subarea of a conference. In general, our opinions are more or less fair depending on our true competences. Thus, CJ should consider this dimension as many existing trust models do [Osman et al., 2013; Sierra and Debenham, 2006]. The inclusion of a semantic dimension on trust and reputation requires defining an ontology of the domain and semantic distances between the elements in the vocabulary. This represents no technical problem and will basically increase the complexity of the computation proportionally to the granularity of the vocabulary.

Finally, in the context of assessing employees' performance, malicious employees may collude to artificially overrate their works. Eigentrust has extensions that are robust against this collusion and can be used as an improvement of CJ [Kamvar et al., 2003].

Chapter 7

Conclusions and Future Work

In this dissertation we tackle several problems relevant to people management in organisations. In particular, we contribute by proposing a people management workflow that integrates team composition techniques, employee fair evaluation and task performance evaluation.

First, we review the most prominent tools to measure individuals' attributes, as these measures are necessary inputs for team composition processes. In particular, we describe the dominant approaches in Organisational Psychology, Industrial Psychology and Human Resources and summarise they main findings to measure individual personality and competences. The most popular personality model is the Five Factor Model (FFM), however, to the best of our knowledge, there are no clear instructions in the literature for team composition including all FFM personality traits. Douglas J. Wilde with his Post-Jungian Personality Theory proposes a team composition model that is based on a modified MBTI questionnaire. The theory is not thoroughly tested, but his research results are promising as his twenty years study gave evidence of the method effectiveness. Regarding competences, organisations are in need to develop their own competence models that can be used across all Human Resources processes including team composition. An organisation can develop its model from scratch or it can use one of the models existing in the literature. We chose to obtain this information from the Multiple Intelligences Theory of Howard Gardner that specifies eight different intelligences (that can be treated as competences). We use a quick and relatively simple method to obtain information of the employees' competences.

Second, we review the literature on team composition and formation from both the organisational psychology and computer science perspectives and we explore the connection between individuals' attributes and team performance as well as the cross fertilization opportunities between those fields. We use our findings to propose a model to predict team performance given a task and based on the individuals' attributes (i.e. competences, personality and gender). We define the Synergistic Team Composition Problem (STCP) as the problem of finding a team partition constrained by size so that each team, and the whole partition of agents into teams, is balanced in terms of individuals' attributes. We propose two different algorithms to solve this problem: an optimal algorithm called STCPSolver that is effective for small instances of the problem, and an approximate algorithm called SynTeam that provides high-quality, but not necessarily optimal solutions.

Third, we devise an algorithm to fairly evaluate individuals' and teams' outcomes once tasks are performed. In particular, we want to diminish the importance of biases in the evaluation process by allowing employees to assess the fairness of appraisals.

In this chapter we summarize and discuss the research work presented in this thesis. In detail, in Section 7.1 we discuss how we have addressed and answered the questions stated in the Introduction and in Section 7.2 we present and discuss some challenging lines for future work.

7.1 Lesson Learned

In this section we discuss the answers to the open questions set up in the Introduction (See Chapter 1). Before responding to each question, we state them again as a reminder.

- ***Question 1:*** *Are there unexplored cross-fertilization ideas between the Computer Science and the Organisational Psychology fields when it comes to team composition and formation?*

We answer ***Question 1*** in Chapter 2, where we analyse the relevant Computer Science (CS) and Organisational Psychology (OP) literature. We elaborate further in Chapter 3 by providing methods from Organisational Psychology and Human Resources to measure individuals' attributes.

CS and OP have followed rather disparate approaches when it comes to team composition and team formation. First, individuals' capacities are far wider concepts in the OP literature as they include competences, experience, gender or age, while CS focuses typically on skills represented as binary attributes of agents (i.e. an agent either has a skill or not). Moreover, while in OP the individuals' capabilities are assumed to be dynamic (i.e. lifelong learning), software agents capabilities are assumed to be static and only the behaviour model may change with agents' interactions. Second, the majority of CS approaches assume that the joint capabilities of agents in a team are enough to solve a given task. However, the researchers in OP recognize also other factors as important when composing and forming a team, such as the motivation of individuals and the task context. They also show that the motivation characteristics predict more accurately the performance of a team than the other factors. Regarding OP research gaps, it lacks a mapping between cognitive ability of individuals and task types (an input in CS models) which complicates team composition. Third, in

OP the performance is assessed from two perspectives: objective and subjective, while CS only considers objective measures. According to OP, team performance cannot be assessed by the time spent to perform a task, by comparing costs or by counting the number of right answers as it would ignore some important subjective reasons. Instead, OP analyzes possible causes of failure, such as an excessive amount of work needed to execute the task given the size of the team or the lack of motivation of team members. Fourth, typically OP analyzes only complex and realistic scenarios as humans have memory and improve their capabilities with every task. On the contrary, since in CS agents can be modeled depending on the needs, researchers can study different settings depending on the dynamics of task arrival (one task or many, one instant of time or several). Researchers in CS use these complex scenarios to let agents build their beliefs based on past experiences and compose new teams according to these learned beliefs.

There are also some similarities between both fields. First, when modeling individuals' attributes, there are two main approaches. In CS, either we base on extensive a-priori information about individuals or we allow individuals to learn their teammates' attributes. Similarly, in OP a number of tests are proposed to acquire a-priori information about teammates (see Chapter 3 for details). Also, OP research allows to learn individual attributes from their repeated interactions. Second, to maximize team performance, one of the crucial findings in both OP and CS is that team members have to be heterogeneous. Third, when it comes to the task execution, both OP and CS focus rather on individuals' attributes required to perform a task than on a detailed planning of the task execution.

Finally, based on the explored range of concepts and issues concerning team composition and formation, we formulated several new research questions for the field. In particular, we identified the following unexplored cross-fertilisation ideas for further research in OP:

1. **Team Composition Automation.** The OP literature has mainly focused on empirically investigating the factors that influence team performance to develop heuristics that help organisations handcraft their teams. OP has disregarded the algorithmic results developed by computer scientists to automate team composition and formation. There is a need in Organisational Psychology research to incorporate models developed by CS.

2. **Matching between task types and team types connected with team performance.** There is a need to understand the correlation between task type and team type and the exact influence on team performance. Currently Organisational Psychology focuses only on classification of team types and task types. We believe it would be beneficial to know the exact correlation between team and task type based on team performance.

3. **Exploring simple team scenarios.** Since in CS agents can be engineered depending on the needs (i.e. agents can be designed with different

attributes, such as personality or memory, depending on the whole system design), researchers can study different settings depending on the dynamics of task arrival. This technology could be used in Organisational Psychology research to understand more exact relationships between individuals' attributes and team performance. Currently, OP research studies only complex scenarios.

Also, we identified the following unexplored cross-fertilisation ideas for further research in CS:

1. **Exploring more complex team models.** A goal of OP is to improve organisational performance by placing the right people in the right jobs, thus enhancing the fit between the individual and the organisation. Research findings from the OP literature have much potential for MAS heuristics (such as team diversity [Mathieu et al., 2008], team size [Mao et al., 2016] or context [Guzzo and Dickson, 1996]). According to OP research, in order to carry out highly interdependent tasks, all team members should possess coordination skills (maturity) and some of them the capacity to take decisions (diversity). Also, the greater the uncertainty and interdependence of task types, the more diverse the competences for team members to cope with complexity. However, if the team is overqualified for the task to perform, the motivation of team members decreases and the quality of the outcome is lower or the task is not completed at all. All these dependencies have been studied extensively by OP research, but they are ignored in CS.

2. **Exploring more complex team members.** In OP, the most important capacity of team members that is related to team performance is their cognitive ability. It is a much wider concept than the notion of capacity in multiagent systems, since beyond skills, widely used by MAS research, it contains many other attributes such as experience, competences, age, or even gender. While some of the human attributes may not make sense in an agent context (like age or gender), some do (such as cognitive abilities, lifelong learning or behavioral model). Also, there is a need to include more sophisticated models for agent capabilities, such as graded capabilities instead of binary ones. Richer agent models would allow the CS field to further benefit from OP findings for team composition and formation. Additionally, the majority of CS models assume that competences are a fixed attribute of each agent. OP indicates that human capabilities are necessarily dynamic (evolve along time) so that teams can successfully improve their performance to solve tasks in realistic dynamic real-world scenarios and in a variety of contexts. The dynamics of competences through learning and experience and human cultural values could be used by MAS research to program adaptive agents, specially when interacting in mixed teams involving humans.

3. **Evaluating team performance.** From an OP perspective, team performance cannot be assessed by simply measuring how long it takes for

a group to finish a certain task or by counting the number of right answers to predefined clear questions, which is a common approach in CS. OP rather analyzes joint team objectives and the team composition and formation setting (such as unrealistic deadlines, the number of individuals in a team, the level of stress in a team or the quality of the outcome). Also, OP focuses on the inner development of team members and analyses the quality of human resources in a team, that is, motivation, satisfaction, commitment, illness or stress rate [Quijano et al., 2008]. When evaluating team performance, Computer Science research should take into account team objectives, task dependencies, the feasibility of the task, etc.

4. **Including agent motivation.** OP research highlights motivation as an important factor for team performance [Hackman, 1990]. The majority of the MAS literature on team composition and teamwork assumes that agents always behave according to their capabilities and knowledge. While in MAS research it is shown that motivation increases by introducing competition mechanisms (like in crowdsourcing teams, [Rokicki et al., 2015]), or by giving agents freedom when selecting their collaborators (like in ad-hoc teams, [Agmon et al., 2014]), these are only early attempts to include agents' motivation as an important factor for team performance.

5. **Exploiting the context.** OP research results suggest that context plays an important role in the performance of teams, [Guzzo and Dickson, 1996; Hackman, 1990; Terveen and McDonald, 2005]. Although, to the best of our knowledge, there are only a few works in CS that would recognize context as an important factor, besides the social and geographical context considered in some papers. There is a need to perform further research on how to computationally model the context within team composition and team formation to build better performing agent teams.

We incorporated some of these ideas in our work. In particular, in our synergistic team composition model, we give preference to the teams that are diverse with respect to individuals' attributes. We also include graded competences instead of binary ones. Finally, we acknowledge the importance of gender balance when composing synergistic teams.

- ***Question 2:*** *Can we predict a single team performance better than experts?*

To answer this question, in Chapter 4 we built a model called the Synergistic Team Composition Model (STCM) to predict performance of a single team given a task taking into account individuals' competences, their personality and their gender. In our experiments, we aimed at checking if using the given individuals' attributes, we were able to predict team performance better than experts. With this purpose, we performed the experiments in an educational scenario. In current school practice, teachers handcraft teams based on expert knowledge about students, their competences, background and social situation.

We decided to compare our automated team composition model with the team composition performed by teachers. We used data of 48 students from a state school “Institut Torras i Bages” in L'Hospitalet de Llobregat, near Barcelona. We generated several team performance rankings using the evaluation values obtained through different methods. First, we generated a ranking based on actual team performance, namely the *base ranking* to compare against. Second, we generated a ranking based on experts' evaluations about the expected performance of teams. Finally, we generated several rankings using synergistic values with varying congeniality and proficiency trade-offs. In particular, we wanted to observe how the rankings changed when increasing the importance of competences. Henceforth, we compared the teachers' and STCM rankings with the actual performance ranking using the generalised standardized Kendall Tau distance (see Section 2.1 for background of this method). We observed that when competences were not included in the STCM ranking, both methods had comparable errors when ranking teams. When we increased the importance of competences, we noticed the improvement of STCM rankings. We obtained the best estimation for $\lambda = 0.8$ for the considered task (creative but also requiring a high level of competences). In other words, based on this experiment, for a creative and competence demanding task the ratio between proficiency and congeniality should be equal to $4 : 1$. We observed that when having this proportion, rankings generated by STCM were very similar to the actual performance ranking. That is, it had only 15% of pairwise disagreements, while teacher's ranking had 28% of pairwise disagreements. These results answer Question 2. We are able to predict the performance better than experts by using the STCM model that takes into account individuals' competences, personality and gender.

- ***Question 3:*** *Is there a method to divide an organisation so that all teams work better than the teams composed by experts?*

In order to answer this question, in Chapter 5 we considered the Synergistic Team Composition Problem (STCP) that is the problem of splitting a set of employees into teams of even size so that each team (and the whole partition of employees into teams) is balanced in terms of competences, personality and gender. We proposed an optimal and an approximate solution to this problem. First, we discussed an algorithm to optimally solve the STCP called STCPSolver. When we noticed that the algorithm is only effective for small instances of the problem, we developed SynTeam, a greedy algorithm for partitioning groups of humans into proficient, gender, psychologically and size balanced teams, which yields a good, but not necessarily optimum solution. Our computational evaluation shows that SynTeam outperforms STCPSolver when the number of employees is large and for big team sizes. Moreover, the SynTeam algorithm provides approximate solutions with good quality guarantees (i.e., up to 87%).

Next, we performed two different experiments in an education scenario with a total of 252 students to show the effectiveness of our method in real-life scenarios. In both experiments we divided each classroom into two halves: one half was divided into teams using SynTeam and the other half using the usual teacher method (see Section Chapter 4.5.1 for the detailed description). In the first experiment we composed teams of size three (31 teams in total) and we asked them to create a tourist brochure of their city. The work was assessed by a panel of teachers. The relative improvement was equal to 29.2%. In the second experiment we composed teams of size two (75 teams in total) and each team had to program a game in Scratch. The work was assessed by an independent Scratch expert who did not know the source of the teaming. In this experiment the relative improvement was equal to 25.3% and it was statistically significant (p-value = 0.04).

These results answer question 3. Using either STCPSolver that composes optimal teams in terms of synergistic values or SynTeam that gives us a high-quality approximate solution, we are able to divide a set of individuals that perform better than teams composed by a tutor that knows the students — their background, competences, social and cognitive capabilities. Notice also that these results further confirm the effectiveness of the STCM model, and thus, strengthen the results presented in Chapter 4.

- ***Question 4:*** *Can we diminish the importance of biases when assessing individual and team performance?*

In order to answer ***Question 4*** in Chapter 6 we were interested in designing an assessment method that was able to identify incompetent reviewers and lower their importance in the contribution to a final performance rating. Henceforth, we proposed a new ranking algorithm called Collaborative Judgment (CJ) that takes into account peer opinions of employees as well as peer judgments over those opinions.

The algorithm can be used to measure competences of employees as well as team performance. For simplicity reasons, we applied CJ to the use case of scientific paper assessment and we validated it over simulated data. The experiment went as follows. First, we assumed that each paper has a “true quality” value that informs us on how good it is. Second, we modelled two kinds of peers: (1) good reviewers provide fair opinions and fair judgments, (2) bad reviewers when giving opinions stay mostly in the central area of the evaluation space as they cannot assess true paper quality (since they are incompetent), but provide reasonable judgments as they can evaluate opinions of others as being informative or not. Finally, we compared CJ with the Self-Assessment Weighted Algorithm (SAWA), which is the standard algorithm used in Conference Management Systems. In contrast to SAWA that treats all reviewers equally, CJ is designed to detect biased reviewers and diminish the importance of their opinions using a computed reputation measure. We were interested in analysing the accuracy of opinions and generated rankings based on those opinions as well as in assessing the robustness against bad reviewers. When it comes to accuracy, we observed that CJ’s gains become larger than 20% and statistically significant for percentages of good reviewers between 20% and 80%. This shows that CJ is able to label bad reviewers accurately and diminish their contribution in a final performance rating. Similarly, we observe that rankings generated by CJ with respect to

SAWA are closer to the true ranking and the improvement asymptotically grows to 42% for 80% of good reviewers. These results show that rankings produced by this new algorithm (under (reasonable) assumptions on reviewer behaviour) improve the self-assessment weighted Average. Finally, when it comes to robustness, we observe that the algorithm outperforms SAWA by 10% when good reviewers are frequently mistaken. When improving the quality of the good reviewer behaviour, the improvements grow up to 51%. These results show that CJ is able to diminish the importance of biases when assessing individual and team performance.

7.2 Future Work

On top of the general ideas for further work on CS and OP team composition listed in the previous section (see Section 7.1, in this section we describe several open issues that raise from the research presented in this thesis, and the corresponding lines of future work that could address them.

In particular, the research introduced in this thesis opens several paths to future developments. We organise these ideas around the notions introduced in the team composition model.

7.2.1 Tasks

- **Exploring a set of different tasks.** In this thesis when composing teams we consider a scenario where we look for teams for multiple copies of one complex task. This scenario is useful for many organisational settings, i.e. in departments such as software engineering or sales departments where there is a need of dividing employees into teams so that each team works on the very same task type. However, we believe that generalizing the problem to finding a partition of teams for a set of different tasks would be useful for organisations. This requires a general version of the current team composition problem. There is a need to introduce tasks that differ in their character and modify the proposed model as well as the algorithms so that each composed team is responsible for a different task.

- **Automating task definition.** In the current solution defining a task is a manual process that consists of defining the number of employees required as well as a task type (i.e. the importance of proficiency in a task type and the list of competence requirements). As part of future work, we could use the task classifications proposed by the organisational psychology literature (see Section 2.5.2 for a detailed description) to cluster tasks that were defined in the past (and hence exist in the system). We could use that clustering to help project managers in defining new tasks' characteristics by searching through *similar* tasks that were once performed. It would also be interesting to learn some parts of task characteristics based on other characteristics. For instance, if we had the competence requirements defined we could come up with *similarity* measures to compare tasks that were performed in the past and based on that recommend other characteristics, such as proficiency importance or team size. Similarly, there may be some highly desired competences that are associated with particular proficiency levels. For example, if we lack experts, we might want to have lower level of proficiency and require some "general" competences such as 'teamwork'. To do this we could explore the list of competences that appear frequently for a given proficiency importance. Finally, given a task and history of teams' performance for similar tasks we could recommend additional competences that members of successful teams had.

- **Recommending task alterations.** When composing a team for a task, we could set a minimum threshold for the team's synergistic value to inform us if the team may have difficulties performing the task. Based on that threshold, a system could recommend changes in the task. For instance, it may happen that the number of required team members is either too small to cover all required competences or too high and some competences are covered multiple times (which may be desirable in some tasks, although not necessarily in others). Hence, a system could recommend potential changes in team size to improve the synergistic value of a team. Similarly, a system could inform us when the level of a particular competence is too low. Finding the thresholds for synergistic values as well as competence levels poses a question for future work.

7.2.2 Employees

- **Adding other attributes.** In this thesis we followed the most prominent recommendations of organisational psychology when it comes to individuals' attributes that influence team performance (i.e. competences / cognitive ability, personality, gender). However, there are also other attributes that were not included in this work, such as motivation, race, age, etc. In future work, we could explore further which attributes to add and how to extend the synergistic model to include them.

- **Exploring Different Personality Tests.** In our research, we composed a model for two different personality tests, i.e. Myers-Briggs Type Indicator that is widely accepted by organisations and the Post-Jungian Personality Theory as a novel and interesting for research method (see Section 3.1 for details). As future work, we would like to explore further findings of the organisational psychology literature to come up with distinct congeniality measures when using different questionnaires. While organisational psychology research on some personality questionnaires shows mixed results when it comes to the relationship between team members personalities and team performance (e.g. the Belbin Theory), research on the Five Factor Model (FFM) strongly suggests a correlation between team members personalities and team performance. However, current organisational psychology studies test each FFM trait one-by-one (rather than as a combination of several traits) and suggest that there are some employees that are simply better than the others (see Section 3.1 for a detailed description). To the best of our knowledge, there were no attempts to examine the FFM model in terms of team member personality configuration. It would be worth exploring ways to compose a model where all employees from an organisation are included. Nonetheless, this would require a thorough study from the organisational psychology field.

- **Introducing a notion of workload.** In real-life scenarios each employee is hired for a particular amount of hours. Therefore, as future work, we would like to explore the notion of workload. That is, we could add for each employee a number in $[0, 1]$ meaning the percentage of employee's time that is already planned. This workload could be used to decrease the size of the problem by filtering a set of employees to exclude the ones that are not available at a time.

- **Direct users' feedback.** We should not assume that an "optimal" AI system is always correct as any model has its limitations. Hence, the system should also take into account the user opinion by allowing them to comment on composed teams. We could introduce it to the model by adding a set of constraints and preferences. For instance, Ana cannot be in the same team with José as they used to have a romantic relationship. It is very difficult to foresee all possible relationships and gather information for those in an automatic way. However, accepting these constraints as part of an algorithm input would be beneficial for many real-life settings. It could be introduced either directly in the model by penalising the utility of teams that violate the constraint or in case of using the optimal solver, when defining the problem.

- **Indirect employees' feedback.** Instead of hard-coding the constraints, as mentioned above, for the system, we could learn them. That is, we could build a graph of peer synergies based on the history of team performance and include them as one more parameter of the synergistic model. This way we could actually learn which employees work best together and use this information to compose even better performing teams.

7.2.3 Teams

- **Finding partial solutions.** The algorithms proposed in this thesis find a partition of synergistic teams. However, in real-life we might face changes in the initial set of employees due to internal or external migration, i.e. absence or inclusion of some employees. Shuffling all employees between teams, especially when teamwork has already started can be problematic. Therefore, we believe that future work could explore methods to compose teams that are at least as synergistic as required while shuffling as few employees as possible.

- **Selecting a number of best teams.** Selecting a suitable set of employees for projects (tasks) is one of the most frequent problems in organisations. More formally, given a list of employees A, we must select the most synergistic team or a list of r most synergistic teams for given task $t = \langle \tau, m \rangle$. To do so, we need to list all possible combinations of employees A into teams of size m, calculate synergistic values for them and choose r teams with the highest synergistic values.

- **Internal and External Recruitment.** Many organisations do not want to compose an entirely new team every time a new complex task arrives. They rather prefer to have stable teams and just add some team members depending on the needs of a task and the level of knowledge of the current team. Therefore, our model can also be used to select a team member whose *synergistic added value* (SAV) to a given team is the highest. A synergistic added value is defined as a relative added value such that:

$$SAV(K, a) = \left(\frac{s(K \cup \{a\})}{s(K)} - 1 \right) \cdot 100\% \tag{7.1}$$

  To find the most suitable team member we must calculate the SAV value for all candidates $a \in A_{cand}$ and select the one whose value is maximal. Hence, it poses a problem of finding $a^* = \arg\max_{a \in A_{cand}} SAV(K, a)$ where A_{cand} is the set of all possible employees that we can add to a given team.

- **A team member removal.** Similarly to the recruitment problem (i.e. adding team members), we might want to reduce the size of a team (i.e. subtract team members). Here, the problem consists of selecting the most redundant team member. To do this we need to compute for all team members $a \in K$ SAV value as following:

$$SAV(K, a) = \left(\frac{s(K)}{s(K \setminus \{a\})} - 1 \right) \cdot 100\% \tag{7.2}$$

  and select an employee whose SAV value is minimal: $a^* = \arg\min_{a \in K} SAV(K, a)$

- **Task alteration recommendation.** Given a team and a task, we can automatically discover redundant team members and recommend their removal. In detail, we can calculate a SAV value for each team member given a task and report team members whose value is lower than zero. This means that either the employee is not as experienced as required, his personality is not compatible with other team members or simply that the team is already sufficiently synergistic for a given task type.

- **Predicting team performance.** In organisations it is fundamental to predict whether a team is going to be successful or fail in problem solving. Henceforth, given several teams that are already composed, we might want to *compare them*. For instance, given a task we might want to compare teams composed for this task with the ones that performed the same task type in the past. To do this, we can calculate the synergistic value for each team and check if the value is at least as high as that of a team in the past. We can also use proficient and congenial values to have the most informed comparison. However, establishing the exact function of team *similarity* poses a separate problem for team research.

7.2.4 Assignments

- **Exploring different assignments.** In the solution proposed in Chapters 4 and 5 we treat an assignment problem as an optimisation problem where we want to have each competence assigned to at least one employee and each employee assigned to at least one competence so that the total *cost* of the assignment is minimal (in terms of under- and over-proficiency). Defining the assignment as the above optimisation problem ensures that all employees are engaged (as each is assigned to at least one competence) and are kept challenged by appointing each competence to an employee whose competence level is the closest to the required one[1]. However, depending on the objectives of team composition, we could define some other assignments. For instance, as mentioned in Subsection 4.3.2, if the purpose of composing (a) team(s) is co-operative learning, we might want to assign exactly two employees per competence — one over-proficient serving as a teacher and one under-proficient acting as a student. We believe that this poses an interesting question to consider in future work. That is, we would like to consider how to define and build the most appropriate assignment given the goal of team composition.

- **Exploring Parallel Computing.** There is a need to explore the recent developments in parallel computing to compute all the synergistic values in parallel. This could be achieved by solving all the minimum cost flow problems in parallel [Sakharnykh and Braun, 2017].

[1] We could assign the most proficient employee to each competence, however as discussed in the beginning of Section 4.4.1 this could cause a drop in employee motivation.

Appendices

Appendix A

Intelligences Test

In this Appendix we present the Multiple Intelligences test developed by Rice ([Rice, 2013]) used in our experiments in schools to measure students' intelligences (see details of this theory in Section 3.2 and details of the experiment in Chapter 5). We provide two versions of the test, i.e. English original version and Spanish translation. During our experiments students we asked students to fill in the test in Spanish.

There are five questions per each intelligence. We are using a five-level Likert item [Likert, 1932] as a model for possible answers for each question, that is:

- Totally disagree / Totalmente en desacuerdo (0),
- Disagree / En desacuerdo (0.25),
- Neutral / Ni de acuerdo ni en desacuerdo (0.5),
- Agree / De acuerdo (0.75),
- Totally Agree / Totalmente de acuerdo (1).

The values in the brackets show the quantitative values associated with the semantics of each answer.

The test of Intelligences in English goes as follows:

- **Naturalist-Environmental**
  1. I feel at home outdoors and in natural surroundings.
  2. Taking care of the environment is a high priority.
  3. Factual Studies and social studies information gives me quality enjoyment time.
  4. I relate well to animals and enjoy responsibility of caring for them.
  5. I am sensitive to the sights, sounds, and feel of things around me.
- **Logical-Mathematical**
  1. I can add or multiply quickly in my head.

2. I like to work with calculators.
3. I like to play number and strategy games.
4. I can see patterns and relationship between numbers quickly and easily.
5. I like to work with numbers and figures.

- **Verbal-Linguistic**

1. It is easy for me to say what I think in an argument or debate.
2. I enjoy a good lecture, speech, or debate.
3. I am irritated when I hear an argument or statement that sounds illogical.
4. I enjoy writing detailed letters to friends.
5. I am good at findings the fine points of word meanings.

- **Visual-Spatial**

1. I would rather draw a map than give someone verbal directions.
2. I always know North from South no matter where I am.
3. I always understand the directions that come with new gadgets and appliances.
4. I can look at an object one way and see it turned sideways or backwards just as easily.
5. Just looking at shapes of buildings and structures is pleasurable to me.

- **Bodily-Kinesthetic**

1. I pick up new dance steps quickly.
2. Learning to ride a bike or skate was easy.
3. My sense of balance and coordination is good.
4. I enjoy building models and sculpters.
5. I am good at athletics.

- **Musical- Rhythmical**

1. I can plan or used to play a musical instrument.
2. I can associate music with my moods.
3. Life seems empty without music.
4. I like to hum, whistle, and sing in the shower or when I am alone.
5. I often connect a piece of music with some event in my life.

- **Interpersonal**

1. I like to gather together groups of people for parties of special events.
2. I have a good sense of what other people think of me.
3. I can convince other people to follow my plans.
4. I am sensitive to the expressions on other people's faces.

5. I am sensitive to the moods of others.

- **Intrapersonal**

  1. If I am angry or happy, I usually know exactly why.
  2. I can help a friend sort out strong feelings because I successfully dealt with similar feelings myself.
  3. I like to sit quietly and reflect on my inner feelings.
  4. I am usually aware of the expression on my face.
  5. I stay 'in touch" with my moods. I have no trouble identifying them.

The test of Intelligences in Spanish goes as follows:

- **Naturalista**

  1. Me siento como en casa al aire libre y en un entorno natural.
  2. Cuidar el medioambiente es una gran prioridad.
  3. Disfruto con la información de estudios experimentales y sociales.
  4. Me relaciono bien con los animales y disfruto con la responsabilidad de cuidarlos.
  5. Soy sensible a la imagenes, sonidos y sensaciones de las cosas que me rodean.

- **Logical-Mathematical**

  1. Puedo sumar o multiplicar mentalmente con mucha rapidez.
  2. Me gusta trabajar con calculadoras y computadores.
  3. Me gusta resolver rompecabezas y entretenerme con juegos de strategia.
  4. Con frecuencia veo configuraciones y relaciones entre números con más rapidez y facilidad que otros.
  5. Me gusta trabajar con números y figuras.

- **Verbal-Linguistic**

  1. No me es difícil decir lo que pienso en el curso de una discusión o debate.
  2. Disfruto de una buena charla, discurso o sermón.
  3. Me pongo nervioso(a) cuando oigo una discusión o una afirmación que parece ilógica
  4. Me gusta escribir cartas detalladas a mis amigos.
  5. Tengo agudeza para encontrar el significado de las palabras.

- **Visual-Spatial**

  1. Prefiero hacer un mapa a explicar a alguien como llegar a un sitio.
  2. Siempre distingo el norte del sur, esté donde esté.

3. Siempre entiendo los gráficos que vienen en las instrucciones de equipos o instrumentos.
4. Puedo distinguir un objeto desde diferentes puntos de vista.
5. Con sólo mirar la forma de construcciones y estructuras me siento a gusto.

- **Bodily-Kinesthetic**
  1. Aprendo rápido a bailar un baile nuevo.
  2. Me fue fácil aprender a ir en bicicleta (o a patinar).
  3. Tengo buen sentido del equilibrio y coordinación.
  4. Me gusta construir modelos (o hacer esculturas).
  5. Soy bueno(a) para el atletismo.
- **Musical- Rhythmical**
  1. Sé tocar o antes sabía tocar un instrumento musical.
  2. Asocio música con mis estados de ánimo.
  3. La vida me parece vacía sin música.
  4. Me gusta tararear, silbar y cantar en la ducha o cuando estoy solo(a).
  5. Con frecuencia hago la conexión entre una pieza de música y algún evento de mi vida.
- **Interpersonal**
  1. Me gusta reunir grupos de personas en una fiesta o en un evento especial.
  2. Me doy cuenta bastante bien de lo que otros piensan de mí.
  3. Soy capaz de convencer a otros para que sigan mis planes
  4. Me doy cuenta de las expresiones en la cara de otras personas.
  5. Me doy cuenta de los estados de ánimo de otros.
- **Intrapersonal**
  1. Si estoy enfadado(a) o contento(a) generalmente sé exactamente porqué.
  2. Puedo ayudar a un(a) amigo(a) a manejar sus sentimientos porque yo lo pude hacer antes con sentimientos parecidos.
  3. Me gusta sentarme silenciosamente y reflexionar sobre mis sentimientos íntimos.
  4. Generalmente me doy cuenta de la expresión que tengo en la cara.
  5. Me mantengo “en contacto” con mis estados de ánimo. No me cuesta identificarlos

Appendix B

Personality Test

In this Appendix we present the Post-Jungian Personality test developed by Wilde ([Wilde, 2013]) used in our experiments in schools to measure students' personalities (see details of this theory in Section 3.1 and details of the experiment in Chapter 5). We provide two versions of the test, i.e. English original version and Spanish translation. During our experiments students we asked students to fill in the test in Spanish.

The original personality test in English is as following:

- **Sensing / Intuition**

You prefer the:	(s) specific	(n) abstract	(0) Indifferent
You prefer:	(s) investigate	(n) speculate	(0) Indifferent
You are more:	(s) practical	(n) conceptual	(0) Indifferent
You are more:	(s) practical	(n) theoretical	(0) Indifferent
You prefer the:	(s) traditional	(n) novel	(0) Indifferent

- **Thinking / Feeling**

You prefer:	(t) logic	(f) empathy	(0) Indifferent
You are more:	(t) honest	(f) diplomatic	(0) Indifferent
You are more:	(t) curious	(f) accomodating	(0) Indifferent
You are more:	(t) skeptical	(f) tolerant	(0) Indifferent
Judges should be:	(t) impartial	(f) merciful	(0) Indifferent

- **Extrovert / Introvert**

You are more:	(e) sociable	(i) reserved	(0) Indifferent
You are more:	(e) expressive	(i) content	(0) Indifferent
You prefer:	(e) groups	(i) individuals	(0) Indifferent
You learn better by	(e) listening	(i) reading	(0) Indifferent
You are more:	(e) talkative	(i) quiet	(0) Indifferent

- **Perceiving / Judging**

You are more:	(p) informal	(j) systematic	(0) Indifferent
You prefer things:	(p) open-ended	(j) planned	(0) Indifferent
You work better:	(p) pressured	(j) unpressured	(0) Indifferent
You prefer:	(p) variety	(j) routine	(0) Indifferent
You are more:	(p) improviser	(j) methodical	(0) Indifferent

The personality test used in our study (in Spanish) is as following:

- **Sensing / Intuition**

Prefieres:	(s) lo concreto	(n) lo abstracto	(0) Indifferent
Prefieres:	(s) investigar	(n) specular	(0) Indifferent
Eres más:	(s) práctico	(n) conceptual	(0) Indifferent
Eres más:	(s) práctico	(n) teórico	(0) Indifferent
Prefieres:	(s) lo tradicional	(n) lo nuevo	(0) Indifferent

- **Thinking / Feeling**

Prefieres:	(t) lógica	(f) empatía	(0) Indifferent
Eres más:	(t) honesto	(f) diplomt́ico	(0) Indifferent
Eres más:	(t) curioso	(f) accomodado	(0) Indifferent
Eres más:	(t) escéptico	(f) tolerante	(0) Indifferent
Los jueces deben ser	(t) imparciales	(f) compasivos	(0) Indifferent

- **Extrovert / Introvert**

Eres más:	(e) sociable	(i) reservado	(0) Indifferent
Eres más:	(e) expresivo	(i) contenido	(0) Indifferent
Prefieres:	(e) grupos	(i) individuos	(0) Indifferent
Aprendes mejor:	(e) escuchando	(i) leyendo	(0) Indifferent
Eres más:	(e) hablador	(i) silencioso	(0) Indifferent

- **Perceiving / Judging**

Eres más:	(p) informal	(j) sistemático	(0) Indifferent
Prefieres cosas:	(p) abiertas	(j) planeadas	(0) Indifferent
Trabajas mejor:	(p) Con presión	(j) Sin presión	(0) Indifferent
Prefieres:	(p) variedad	(j) rutina	(0) Indifferent
Eres más:	(p) Improvisador	(j) Metódico	(0) Indifferent

Bibliography

Agmon, N., Barrett, S., and Stone, P. (2014). Modeling uncertainty in leading ad hoc teams. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'14)*.

Ahuja, R., Magnanti, T., and Orlin, J. (1993). *Network flows: theory, algorithms, and applications*. Prentice hall.

Alberola, J. M., Del Val, E., Sanchez-Anguix, V., Palomares, A., and Teruel, M. D. (2016). An artificial intelligence tool for heterogeneous team formation in the classroom. *Knowledge-Based Systems*, 101:1–14.

Alfaro, L. and Shavlovsky, M. (2013). Crowdgrader: Crowdsourcing the evaluation of homework assignments. *Tech. Report 1308.5273, arXiv.org*.

Amigó, I. J. (2018). Aprenentatge cooperatiu. http://www.insjoanamigo.cat/web/attachments/article/118/Treball%20cooperatiu.pdf.

Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., and Leonardi, S. (2012). Online team formation in social networks. In *Proceedings of the 21st International Conference on World Wide Web (WWW'12)*, pages 839–848.

Anderson, N., Ones, D. S., Sinangil, H. K., and Viswesvaran, C. (2001). *Handbook of industrial, work & organizational psychology: Volume 1: Personnel psychology*. Sage.

Anderson, N. and Sleap, S. (2004). An evaluation of gender differences on the belbin team role self-perception inventory. *Journal of Occupational and Organizational Psychology*, 77(3):429–437.

Andrejczuk, E., Rodríguez-Aguilar, J. A., Roig, C., and C., S. (2017). Synergistic team composition (extended abstract). In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'17)*, pages 1463–1465. International Foundation for Autonomous Agents and Multiagent Systems.

Aritzeta, A., Swailes, S., and Senior, B. (2007). Belbin's team role model: Development, validity and applications for team building. *Journal of Management Studies*, 44(1):96–118.

Arnold, J. and Randall, R. (2010). *Work psychology.* Pearson Education Limited., Harlow, England.

Barkley, E. F., Cross, K. P., and Major, C. H. (2014). *Collaborative learning techniques: A handbook for college faculty.* John Wiley & Sons.

Barrett, S., Stone, P., Kraus, S., and Rosenfeld, A. (2013). Teamwork with limited knowledge of teammates. In *Proceedings of the 27th National Conference on Artificial Intelligence (AAAI'13).*

Barth, A. L. and de Beer, W. (2017). *Performance Management Success.* Springer.

Bartol, K. M. (1977). Building synergistic edp teams. In *Proceedings of the fifteenth annual SIGCPR conference*, pages 18–30. ACM.

Bartram, D. and Roe, R. (2005). Definition and assessment of competences in the context of the european diploma in psychology. *European Psychologist*, 10(2):93.

Bashur, R., Hernandez, A., and Peiro, J. (2011). The impact of underemployment on individual and organizational performance. *Underemployment: Psychological, Economic, and Social Challenges içinde (ss. 187-213) New York: Springer.*

Batenburg, R., van Walbeek, W., and in der Maur, W. (2013). Belbin role diversity and team performance: is there a relationship? *Journal of Management Development*, 32(8):901–913.

Bateson, J., Wirtz, J., Burke, E., and Vaughan, C. (2013). When hiring, first test, and then interview. *Harvard Business Review.*

Bear, J. B. and Woolley, A. W. (2011). The role of gender in team collaboration and performance. *Interdisciplinary science reviews*, 36(2):146–153.

Belbin, R. (1993). *Team Roles at Work: A Strategy for Human Resource Management.* Butterworth-Heinemann, Oxford.

Bell, S. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, 92(3):595–615.

Boyle, G. J. (1995). Myers-briggs type indicator (mbti): Some psychometric limitations. *Australian Psychologist*, 30(1):71–74.

Boyle, G. J. (2008). Critique of the five-factor model of personality. *The SAGE handbook of personality theory and assessment*, 1:295–312.

Bradley, J. and Hebert, F. (1997). The effect of personality type on team performance. *Journal of Management Development*, 16:337–353.

Briggs, I. and Myers, P. (1995 [1980]). *Gifts Differing: Understanding Personality Type*. Mountain View, CA: Davies-Black Publishing.

Brodbeck, F. (2003). Team climate for learning in higher education. *Aston Network (Summer)*, pages 4–5.

Broucek, W. G. and Randell, G. (1996). An assessment of the construct validity of the belbin self-perception inventory and observer's assessment from the perspective of the five-factor model. *Journal of Occupational and Organizational Psychology*, 69(4):389–405.

Buckingham, M. (2011). The fatal flaw with 360 surveys. *Harvard Business Review*.

Butchibabu, A., Sparano-Huiban, C., Sonenberg, L., and Shah, J. (2016). Implicit coordination strategies for effective team communication. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(4):595–610.

Campion, M., Medsker, G., and Higgs, A. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel psychology*, 46(4):823–847.

Carley, K. M., Moon, I. C., Schneider, M., and Shigiltchoff, O. (2005). Detailed analysis of factors affecting team success and failure in the america's army game. *Tech Report*.

Carter, P. (2009). *The complete book of intelligence tests: 500 exercises to improve, upgrade and enhance your mind strength*. John Wiley & Sons.

Chalkiadakis, G. and Boutilier, C. (2012). Sequentially optimal repeated coalition formation under uncertainty. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'12)*, volume 24, pages 441–484.

Chamorro-Premuzic, T. (2015). Ace the assessment. *Harvard Business Review. Retrieved from https://hbr. org/2015/07/ace-the-assessment*.

Chan, C. C., Lim, L., and Kuan Keasberry, S. (2003). Examining the linkages between team learning behaviors and team performance. *The Learning Organization*, 10(4):228–236.

Chen, B., Chen, X., Timsina, A., and Soh, L. (2015). Considering agent and task openness in ad hoc team formation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15)*, pages 1861–1862.

Chiocchio, F., Kelloway, E. K., and Hobbs, B. (2015). *The psychology and management of project teams*. Oxford University Press.

Cormen, T. H., Stein, C., Rivest, R. L., and Leiserson, C. E. (2001). *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition.

Costa, P.T., J. and McCrae, R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Odessa, FL: Psychological Assessment Resources.

Crawford, C., Rahaman, Z., and S., S. (2016). Evaluating the efficiency of robust team formation algorithms. *International Workshop on Optimisation in Multi-Agent Systems*.

Davey, L. (2017). If your team agrees on everything, working together is pointless. *Harvard Business Review*.

Deb, T. (2006). *Strategic approach to human resource management*. Atlantic Publishers & Dist.

Delgado Piña, M. I., Romero Martínez, M. A., and Gómez Martínez, L. (2008). Teams in organizations: a review on team effectiveness. *Team Performance Management: An International Journal*, 14(1/2):7–21.

Devine, D. J. (2002). A review and integration of classification systems relevant to teams in organizations. *Group Dynamics: Theory, Research, and Practice*, 6(4):291.

Devine, D. J. and Philips, J. L. (2001). Do smarter teams do better a meta-analysis of cognitive ability and team performance. *Small Group Research*, 32(5):507–532.

Devito, A. J. (1985). Review of myers-briggs type indicator. *The ninth mental measurements yearbook*, 2:1030–1032.

Dubois, D. and Rothwell, W. (2004). *Competency-Based Human Resource Management: Discover a New System for Unleashing the Productive Power of Exemplary Performers*. Nicholas Brealey Publishing.

Duhigg, C. (2016). What google learned from its quest to build the perfect team. *The New York Times Magazine*, 26:2016.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44:350–383.

Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., and Vee, E. (2004). Comparing and aggregating rankings with ties. In *Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '04, pages 47–58, New York, NY, USA. ACM.

Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., and Vee, E. (2006). Comparing partial rankings. *SIAM Journal on Discrete Mathematics*, 20(3):628–648.

Farhangian, M., Purvis, M., Purvis, M., and Savarimuthu, B. (2015a). Modeling the effects of personality on team formation in self-assembly teams. In *Proceedings of the 18th Conference on Principles and Practice of Multi-Agent Systems (PRIMA'15)*, pages 538–546.

Farhangian, M., Purvis, M., Purvis, M., and Savarimuthu, T. B. R. (2015b). Agent-based modeling of resource allocation in software projects based on personality and skill. In *Advances in Social Computing and Multiagent Systems*, pages 130–146. Springer.

Garavan, T. and McCarthy, A. (2008). Collective learning processes and human resource development. *Advances in Developing Human Resources*, 10:451–471.

Gardner, H. and Hatch, T. (1989). Educational implications of the theory of multiple intelligences. *Educational researcher*, 18(8):4–10.

Genter, K., Laue, T., and Stone, P. (2016). Three years of the robocup standard platform league drop-in player competition: Creating and maintaining a large scale ad hoc teamwork robotics competition. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, pages 1–31.

Gibson, C. B. and Kirkman, B. L. (1999). Our past, present, and future in teams: The role of human resource professionals in managing team performance. *Evolving practices in human resource management: Responses to a changing world of work*, pages 90–117.

Goldberg, A. V. and Tarjan, R. E. (1990). Finding minimum-cost circulations by successive approximation. *Mathematics of Operations Research*, 15(3):430–466.

Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216.

Goodwin, G. F., Burke, C. S., Wildman, J. L., and Salas, E. (2009). *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches*. Routledge.

Gusdorf, M. L. (2008). Recruitment and selection: Hiring the right person. *USA: Society for Human Resource Management*.

Guzzo, R. and Dickson, M. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual review of psychology*, 47(1):307–338.

Hackman, J. and Lawler, E. (1971). Employee reactions to job characteristics. *Journal of applied psychology*, 55(3):259–286.

Hackman, J. and Oldham, G. (1975). Development of the job diagnostic survey. *Journal of Applied psychology*, 60(2):159–170.

Hackman, J. R. (1990). *Groups that work (and those that don't): Creating conditions for effective teamwork*. Number 10–H123 in Jossey-Bass management series. Jossey-Bass.

Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Harvard Business Press.

Hang, C.-W. and Singh, M. P. (2012). Generalized framework for personalized recommendations in agent networks. *Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 25(3):475–498.

Hanna, N. and Richards, D. (2015). Do birds of a feather work better together? the impact of virtual agent personality on a shared mental model with humans during collaboration. In *Proceedings of the 3rd International Workshop on Collaborative Online Organizations, COOS 2016, co-located with the 14th International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS 2015, Istanbul, Turkey, May 4, 2015.*, pages 28–37.

Haque, M., Egerstedt, M., and Rahmani, A. (2013). Multilevel coalition formation strategy for suppression of enemy air defenses missions. *Journal of Aerospace Information Systems*, 10(6):287–296.

Hill, R. E. (1975). Interpersonal compatibility and workgroup performance. *The Journal of Applied Behavioral Science*, 11(2):210–219.

Hill, T. and Lewicki, P. (2006). *Statistics: methods and applications: a comprehensive reference for science, industry, and data mining*. StatSoft, Inc.

Hoffman, G. and Breazeal, C. (2004). Collaboration in human-robot teams. In *Proceedings of the 1st AIAA'04 Intelligent Systems Conference*.

Holmes, R. (2013). The unexpectedly high cost of a bad hire. *Hootsuite, https://www. linkedin. com/pulse/20130716151946-2967511-the-high-costs-of-a-bad-hire-andhow-to-avoid-them*.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, pages 324–340.

Horwitz, S. and Horwitz, I. (2007). The effect of team diversity on team outcomes: A meta-analyitic review of team demography. *Journal of Management*, 33(6):987–1015.

IBM (2017). Ibm ilog cplex optimization studio. https://www.ibm.com/us-en/marketplace/ibm-ilog-cplex.

James, L., Choi, C., Ko, C., McNeil, P., Minton, M., and Wright, M. (2008). Collective learning processes and human resource development. *European Journal of Work and Organizational Psychology*, 17:5–32.

Jang, K. L., Livesley, W. J., Angleitner, A., Riemann, R., and Vernon, P. A. (2002). Genetic and environmental influences on the covariance of facets defining the domains of the five-factor model of personality. *Personality and individual Differences*, 33(1):83–101.

Jennings, N., Moreau, L., Nicholson, D., Ramchurn, S., Roberts, S., Rodden, T., and Rogers, A. (2014). Human-agent collectives. *Communications of the ACM*, 57(12):80–88.

Jung, C. (1971). *Psychological types*. Princeton University Press, Princeton.

Kamvar, S. D., Schlosser, M. T., and Garcia-Molina, H. (2003). The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International Conference on World Wide Web (WWW'03)*, pages 640–651, New York, NY, USA. ACM.

Kargar, M., An, A., and Zihayat, M. (2012). Efficient bi-objective team formation in social networks. In *Proceedings of the 2012 European conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD'12)*, pages 483–498, Berlin, Heidelberg. Springer-Verlag.

Katzenbach, J. R. and Smith, D. K. (2015). *The wisdom of teams: Creating the high-performance organization*. Harvard Business Review Press.

Knapp, R. (2010). Collective (team) learning process models: A conceptual review. *Human Resource Development Review*, 9:286–299.

Komaki, J. L. (1997). Behind the scenes: field testing a measure of effectiveness for theater teams. *Brannick, MT, Salas, E., & Prince, C.(E˜ S.), Team Performance Assessment and Measurement: Theory, Methods, and Applications, Mahwah, NJ., Lawrence Erlbaum Associates*, pages 227–25.

Kozlowski, S. and Bell, B. (2013). *Work groups and teams in organizations: Review update*. Cornell University, School of Industrial and Labor Relations.

Kozlowski, S. and Ilgen, D. (2006). Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7:77–28.

Kunesh, T. (2018). Multiple intelligence test for children. http://kids.lovetoknow.com/wiki/Multiple_Intelligence_Test_for_Children.

Laal, M. and Salamati, P. (2012). Lifelong learning; why do we need it? *Procedia-Social and Behavioral Sciences*, 31:399–403.

Lappas, T., Liu, K., and Terzi, E. (2009). Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, pages 467–476, New York, NY, USA. ACM.

Leavitt, A., Keegan, B. C., and Clark, J. (2016). Ping to win?: Non-verbal communication and team performance in competitive online multiplayer games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4337–4350. ACM.

Liemhetcharat, S. and Veloso, M. (2012). Modeling and learning synergy for team formation with heterogeneous agents. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'12)*, pages 365–374.

Liemhetcharat, S. and Veloso, M. (2014). Team formation with learning agents that improve coordination. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'14)*, pages 1531–1532, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Lorenz, J., Rauhut, H., Schweitzer, F., and Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22):9020–9025.

Lu, J. and Zhang, Z. (2012). Understanding the effectiveness of online peer assessment: A path model. *Journal of Educational Computing Research*, 46(3):313–333.

Maghami, M. and Sukthankar, G. (2011). An agent-based simulation for investigating the impact of stereotypes on task-oriented group formation. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pages 252–259. Springer.

Malone, T. W. and Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR)*, 26(1):87–119.

Mao, A., Mason, W., Suri, S., and Watts, D. J. (2016). An experimental study of team size and performance on a complex task. *PLOS ONE*, 11(4):1–22.

Marcolino, L., Jiang, A., and Tambe, A. (2013). Multi-agent team formation: Diversity beats strength? In *Proceedings of the 22nd International Joint Conference (IJCAI'13)*.

Marcolino, L. S., Xu, H., Gerber, D., Kolev, B., Price, S., Pantazis, E., and Tambe, M. (2016). Multi-agent team formation for design problems. In *Coordination, Organizations, Institutions and Norms in Agent Systems XI. Springer-Verlag Lecture Notes in AI.*

Marks, M., Mathieu, J., and Zaccaro, S. (2001). A temporally based framework and taxonomy of team processes. *Academy of management review*, 26(3):356–376.

Martz, W.B., J., Vogel, R., and Nunamaker, J. (1992). Electronic meeting systems: Results from the field. *Decision Support Systems*, 8(2):141–158.

Mathieu, J., Maynard, M., Taylor, S., Gilson, L., and Rudy, T. (2007). An examination of the effects of organizational district and team contexts on team processes and performance. *Journal of Organizational Behavior*, 28:891–910.

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

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., and Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of applied psychology*, 79(4):599.

McGrath, J. E. (1984). *Groups: Interaction and performance*, volume 14. Prentice-Hall Englewood Cliffs, NJ.

Meneses, R. and Navarro, J. (2015). How to improve team effectiveness through group processes: an example in the automotive industry. *Papeles del Psicólogo*, 36(3):224–229.

Mohammed, S. and Angell, L. C. (2003). Personality heterogeneity in teams: Which differences make a difference for team performance? *Small group research*, 34(6):651–677.

Moon, I. C., Carley, K. M., Schneider, M., and Shigiltchoff, O. (2005). Detailed analysis of team movement and communication affecting team performance in the america's army game. *CASOS Technical Report*.

Moos, R. H. and Speisman, J. C. (1962). Group compatibility and productivity. *The Journal of Abnormal and Social Psychology*, 65(3):190.

Mount, M., Barrick, M., and Stewart, G. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human performance*, 11(2-3):145–65.

Myers, I. B., McCaulley, M. H., Quenk, N. L., and Hammer, A, L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*, volume 3. Consulting Psychologists Press Palo Alto, CA.

Nagarajan, V., Marcolino, L. S., and Tambe, M. (2015). Every team deserves a second chance: Identifying when things go wrong (student abstract version). In *Proceedings of the 29th National Conference on Artificial Intelligence (AAAI'15)*, pages 4184–4185.

Nair, R., Tambe, M., and Marsella, S. (2003). Team formation for reformation in multiagent domains like robocuprescue. In Kaminka, G. A., Lima, P. U., and Rojas, R., editors, *RoboCup 2002: Robot Soccer World Cup VI*, volume 2752 of *Lecture Notes in Computer Science*, pages 150–161. Springer Berlin Heidelberg.

Nash, J. (1950). The Bargaining Problem. *Econometrica*, 18(2):155–162.

Navarro, J., Quijano, S., Berger, R., and Meneses, R. (2011). Work-groups in organizations: A basic tool to manage increasing complexity and ambiguity. *Papeles del Psicologo*, 32:17–28.

Noruzi, M. R. and Rahimi, G. R. (2010). Multiple intelligences, a new look to organizational effectiveness. *Journal of Management Research*, 2(2):1.

OEC^2 Solutions, L. (2018). Myers briggs type indicator® (mbti). https://oec2solutions.com/assessment/psychometric-assessment/myers-briggs-type-indicator-mbti/.

Okimoto, T., Schwind, N., Clement, M., Ribeiro, T., Inoue, K., and Marquis, P. (2015). How to form a task-oriented robust team. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15)*, AAMAS '15, pages 395–403. International Foundation for Autonomous Agents and Multiagent Systems.

Olivera, F. and Argote, L. (1999). Organizational learning and new product development: Core processes. *Shared cognition in organizations: The management of knowledge*, pages 297–326.

Orlin, J. B. (1993). A faster strongly polynomial minimum cost flow algorithm. *Operations research*, 41(2):338–350.

Osatuyi, B. J. (2012). *Effects of information importance and distribution on information exchange in team decision making*. New Jersey Institute of Technology.

Osipov, K. and Sukthankar, G. (2012). Amalgacloud: Social network adaptation for human and computational agent team formation. *HUMAN JOURNAL*, 1(2):61–73.

Osman, N., Gutierrez, P., and Sierra, C. (2015). Trustworthy advice. *Knowledge-Based Systems*, 82:41–59.

Osman, N., Sierra, C., McNeill, F., Pane, J., and Debenham, J. K. (2013). Trust and matching algorithms for selecting suitable agents. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):16.

Osman, N., Sierra, C., and Sabater-Mir, J. (2010). Propagation of opinions in structural graphs. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI'10)*, pages 595–600.

Ouye, J. A. (2011). Five trends that are dramatically changing work and the workplace. *Knoll Workplace Research*, pages 1–14.

Oyster, C. (1999). *Groups: A User's Guide*. McGraw-Hill, Boston.

Parker, A. M. and Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, 18(1):1–27.

Partington, D. and Harris, H. (1999). Team role balance and team performance: an empirical study. *Journal of Management Development*, 18(8):694–705.

Peeters, M. A., Van Tuijl, H. F., Rutte, C. G., and Reymen, I. M. (2006). Personality and team performance: a meta-analysis. *European Journal of Personality*, 20(5):377–396.

Peleteiro, A., Burguillo-Rial, J., Luck, M., Arcos, J., and Rodríguez-Aguilar, J. (2015). Using reputation and adaptive coalitions to support collaboration in competitive environments. *Engineering applications of artificial intelligence*, 45:325–338.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., and Koller, D. (2013). Tuned models of peer assessment in moocs. *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*.

Podsakoff, P. M., MacKenzie, S. B., and Ahearne, M. (1997). Moderating effects of goal acceptance on the relationship between group cohesiveness and productivity. *Journal of Applied Psychology*, 82(6):974.

Poropat, A. (2002). The relationship between attributional style, gender and the five-factor model of personality. *Personality and individual differences*, 33(7):1185–1201.

Prewett, M. S., Walvoord, A. A., Stilson, F. R., Rossi, M. E., and Brannick, M. T. (2009). The team personality–team performance relationship revisited: The impact of criterion choice, pattern of workflow, and method of aggregation. *Human Performance*, 22(4):273–296.

Quijano, S., Navarro, J., Yepes, M., Berger, R., and Romeo, M. (2008). Human system audit (hsa) for the analysis of human behaviour in organizations. *Papeles del Psicólogo*, 29(1):92–106.

Rahwan, T., Michalak, T., Elkind, E., Faliszewski, P., Sroka, J., Wooldridge, M., and Jennings, N. (2011a). Constrained coalition formation. In Burgard, W. and Roth, D., editors, *Proceedings of the 25th National Conference on Artificial Intelligence (AAAI'11)*. AAAI Press.

Rahwan, T., Michalak, T. P., Elkind, E., Faliszewski, P., Sroka, J., Wooldridge, M., and Jennings, N. R. (2011b). Constrained coalition formation. In Burgard, W. and Roth, D., editors, *AAAI*. AAAI Press.

Rahwan, T., Michalak, T. P., Wooldridge, M., and Jennings, N. R. (2015). Coalition structure generation: A survey. *Artificial Intelligence*, 229:139–174.

Ramchurn, S., Wu, F., Fischer, J., Reece, S., W, J., Roberts, S., Rodden, T., and Jennings, N. (2016). Human-agent collaboration for disaster response. *Journal of Autonomous Agents and Multi-Agent Systems*, pages 1–30.

Ramchurn, S. D., Farinelli, A., Macarthur, K. S., and Jennings, N. R. (2010). Decentralized coordination in robocup rescue. *Comput. J.*, 53(9):1447–1461.

Ramezan, M. (2011). Intellectual capital and organizational organic structure in knowledge society: How are these concepts related? *International Journal of Information Management*, 31:88–95.

Ramirez-Heller, B., Berger, R., and Brodbeck, F. (2014). Does an adequate team climate for learning predict team effectiveness and innovation potential? a psychometric validation of the team climate questionnaire for learning in an organizational context. *Procedia - Social and Behavioral Sciences*, 114:543–550.

Rangapuram, S., Bühler, T., and Hein, M. (2013). Towards realistic team formation in social networks based on densest subgraphs. In *Proceedings of the 22nd International Conference on World Wide Web (WWW'13)*, pages 1077–1088. ACM.

Rice, L. (2013). *Common Sense Assessment in the Classroom.* Shell Education.

Rico, R., Alcover, C., and Tabernero, C. (2010). Efectividad de los equipos de trabajo, una revision de la ultima decada de investigacion (1999- 2009). *Revista de Psicologia del Trabajo y de las Organizaciones*, 26:47–71.

Rochlin, I., Aumann, Y., Sarne, D., and Golosman, L. (2016). Efficiency and fairness in team search with self-interested agents. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15)*, volume 30, pages 526–552. Springer.

Roe, R. (2002a). Competences-a key towards the integration of theory and practice in work psychology. *Gedrag en Organisatie*, 15(4):203–224.

Roe, R. (2002b). What makes a competent psychologist? *European Psychologist*, 7(3):192.

Rokicki, M., Zerr, S., and Siersdorfer, S. (2015). Groupsourcing: Team competition designs for crowdsourcing. In *Proceedings of the 24th International Conference on World Wide Web (WWW'15)*, pages 906–915.

Rolf, H. L. (2008). *Finite Mathematics, Seventh Edition.* Thomson Learning, Canada.

Rupp, D. E., Hoffman, B. J., and Bischof, D. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management*, 41(4):1244–1273.

Ryan, A., McFarland, L., and SHL, H. B. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52(2):359–392.

Saba (2017). How to set up an effective 360 degree multirater evaluation process. https://www.halogensoftware.com/uk/learn/how-to/how-to-set-up-an-effective-360-degree-multirater-evaluation-process.

Sakharnykh, N. and Braun, H. (2017). Efficient maximum flow algorithms. *GPU Technology Conference*.

Salas, E., Sims, D., and Burke, C. (2005). Is there a "big five" in teamwork? *Small group research*, 36(5):555–599.

Schmidt, F. L. and Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, 124(2):262.

Schutz, W. (1958). *FIRO: A Three Dimensional Theory of Interpersonal Behavior*. Holt, Rinehart, Winston, New York, NY, USA.

Schutz, W. C. (1955). What makes groups productive? *Human Relations*, 8(4):429–465.

Senior, B. (1997). Team roles and team performance: is there 'really'a link? *Journal of occupational and organizational psychology*, 70(3):241–258.

Shaout, A. and Yousif, M. K. (2014). Performance evaluation–methods and techniques survey. *International Journal of Computer and Information Technology*, 3(05):2279–0764.

Shardanand, U. and Maes, P. (1995). Social information filtering: Algorithms for automating “word of mouth”. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*, pages 210–217, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Shaw, M. E. and Webb, J. N. (1982). When compatibility interferes with group effectiveness: Facilitation of learning in small groups. *Small Group Behavior*, 13(4):555–564.

Sierra, C. and Debenham, J. K. (2006). Trust and honour in information-based agency. In *Proceedings of the 5th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)*, pages 1225–1232.

Spradling, M., Goldsmith, J., Liu, X., Dadi, C., and Li, Z. (2013). Roles and teams hedonic game. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT'13)*, pages 351–362. Springer.

Steiner, I. D. (2007). *Group Process and Productivity (Social Psychological Monograph)*. Academic Press Inc.

Stepanyan, K., Mather, R., Jones, H., and Lusuardi, C. (2009). Student engagement with peer assessment: A review of pedagogical design and technologies. *Advances in Web Based Learning–ICWL 2009*, pages 367–375.

Sukthankar, G., Sycara, K., Giampapa, J. A., and Burnett, C. (2009). Communications for agent-based human team support. In *Handbook of Research on Multi-agent Systems: Semantics and Dynamics of Organizational Models*, pages 285–313. IGI Global.

Terveen, L. and McDonald, D. W. (2005). Social matching: A framework and research agenda. *ACM transactions on computer-human interaction (TOCHI)*, 12(3):401–434.

Tesluk, P., Mathieu, J. E., Zaccaro, S. J., and Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. *Team performance assessment and measurement: Theory, methods, and applications*, pages 197–224.

Tetlock, P. E., Peterson, R. S., McGuire, C., Chang, S., and Feld, P. (1992). Assessing political group dynamics: a test of the groupthink model. *Journal of personality and social psychology*, 63(3):403.

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology*, 85(4):723.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3):249–276.

van de Water, H., Ahaus, K., and Rozier, R. (2008). Team roles, team balance and performance. *Journal of Management Development*, 27(5):499–512.

Vosniadou, S. (2003). How children learn. *Successful schooling*, 16.

Walsh, T. (2014). The peerrank method for peer assessment. In Schaub, T., Friedrich, G., and O'Sullivan, B., editors, *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI'14) Including Prestigious Applications of Intelligent Systems (PAIS'14)*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 909–914. IOS Press.

Washington, E. and Griffiths, B. (2015). *Competencies at work: Providing a common language for talent management*. Business Expert Press.

West, M. (2012a). *Effective teamwork: Practical lessons from organizational research*. John Wiley and Sons.

West, M. (2012b). *Effective Teamwork: Practical Lessons Learned from Organizational Research*. Wiley-Blackwell, West Sussex.

White, K. B. (1984). Mis project teams: An investigation of cognitive style implications. *MIS Quarterly*, pages 95–101.

Wilde, D. (2009). *Teamology: The Construction and Organization of Effective Teams*. Springer-Verlag, London.

Wilde, D. (2011). *Jung's Personality Theory Quantified*. Springer-Verlag London.

Wilde, D. (2013). *Post-Jungian Personality Theory for Individuals and Teams*. SYDROSE LP.

Wildman, J. L., Thayer, A. L., Rosen, M. A., Salas, E., Mathieu, J. E., and Rayne, S. R. (2012). Task types and team-level attributes: Synthesis of team classification literature. *Human Resource Development Review*, 11:97–129.

Woolley, A. W., Aggarwal, I., and Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, 24(6):420–424.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, 330(6004):686–688.

Wu, J., Chiclana, F., and Herrera-Viedma, E. (2015a). Trust based consensus model for social network in an incomplete linguistic information context. *Applied Soft Computing*, 35:827–839.

Wu, J., Chiclana, F., and Herrera-Viedma, E. (2015b). Trust based consensus model for social network in an incomplete linguistic information context. *Applied Soft Computing*, 35:827–839.

Yildir, I. (2005). Group motivation and performance indicators in an online team role playing game. *Doctor of Education Thesis, University of Southern California, Southern California*.

Zellmer-Bruhn, M.and Gibson, C. (2006). Multinational organization context: Implications for team learning and performance. *Academy of Management Journal*, 494:501–518.

Zhang, J., Ghorbani, A. A., and Cohen, R. (2007). A familiarity-based trust model for effective selection of sellers in multiagent e-commerce systems. *International Journal of Information Security*, 6(5):333–344.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., and Kraft, M. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of personality and social psychology*, 65(4):757.