ABSTRACT

Title of dissertation: The Effect of Role Specialization And Transactive Memory Systems On Performance in Data Science Teams

Joohee Choi
Doctor of Philosophy, 2020

Dissertation directed by: Professor Yla Tausczik
College of Information Studies

Teamwork is an integral part of data science work. Data science work requires knowledge from many different disciplines including statistics, information visualization, programming, and subject matter knowledge related to a given set of data sets (e.g., politics, education). Data science teams are often formed by individuals who have different areas of knowledge and expertise and, as a result, may take on different functional roles within a team. Due to their distinctive expertise, members in data science teams may take on specialized task roles matching their expertise, and such division of labor could increase coordination cost among team members. As data science work is often open-ended and dynamic by nature, high coordination costs could deteriorate performance in data science teams. In this research, I argued that developing shared cognition on who-knows-what (i.e., transactive memory system, abbreviated as TMS) in data science teams would be beneficial for team performance, especially when the members have specialized roles. I conducted two studies to understand the effect of role specialization and transactive memory systems on

team performance with a goal to identify and test a lever to facilitate transactive memory system in data science teams. I collected data from two consecutive Data Challenge events; Data Challenge is an week-long data science competition hosted annually as a university-wide event. In Study 1, I conducted an observational study by collecting survey data from 74 individuals in 36 teams in Data Challenge 2019. In Study 2, I conducted a field experiment to examine the effectiveness of an experimental intervention designed to facilitate transactive memory system in data science teams by highlighting any inaccuracies in the perceived expertise between members.

The Effect of Role Specialization And Transactive Memory Systems On Performance in Data Science Teams

by

Joohee Choi

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Yla Rebecca Tausczik, Chair/Advisor
Professor Kathryn Bartol, Dean's representative
Professor Brian Butler
Professor Joel Chan
Professor Wayne Lutters

Acknowledgments

I thank my advisor, Dr. Yla Tausczik, for her limitless patience and guidance throughout my doctoral study. I also thank my committee members and the faculties in iSchool who offered invaluable learning opportunities through constructive feedback and encouragement whenever I reached out for one. My friends I met before UMD, at UMD, at conferences, and at various work places, I thank you for the inspiring interactions we had that helped me stay motivated and inquisitive over the years.

I owe my gratitude to my family for providing the sense of a safety net which I can always fall back into whenever I felt overwhelmed. Especially the love and care from my partner, H.K., warrants mention as it has been crucial for completing this dissertation and getting ready for the next step during pandemic. Last but not least, I thank Bono for his relentless whining on a walk that allowed me to take my mind off things at least once a day.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Data science has recently emerged as a new discipline. As the discipline is still actively forming, various definitions of the field coexist. Broadly construed, any activity related to extracting insights from data can be considered data science work [1]. A more conservative view on data science views the field as an extension of statistics, a discipline with a long-standing history [2]. However, it is more realistic to view data science as emerging from a combination of existing fields, with societal changes introducing a new set of challenges, including large scale datasets and abundant unstructured data such as text and image data. These challenges gave birth to a new discipline in its own right, data science [2–4].

Teamwork is an integral part of data science mainly because of the breadth of knowledge required to conduct data science work [5]. Given that data science work forges knowledge from various domains into one project, one may argue that data science teams should be composed of members from diverse functional backgrounds (i.e., educational or professional backgrounds characterized by academic field and job experience). Despite the potential benefit of this diversity, some data science practitioners are more concerned about building data science teams with specialists, each focusing on a distinct role [6,7]. If members in a team exclusively practice their own specialty, it can lead to divisional role structure in teams. As data science work is open-ended and dynamic, with analysts actively forming and updating hypotheses based on analysis results, members have to continuously learn together and stay on the same page to move forward together [6]. Data science practitioners argue that a team of specialists is less ideal than a team of generalists because it usually takes more effort for specialists to coordinate their actions. For example, differences in the use of jargon, interpretations of the same word, and motivation incur increased coordination cost [8].

Coordination in the context of team work is defined as the management of dependencies among tasks, resources, or people [9, 10]. For example, waiting to edit a draft document until an initial draft is completed is due to the dependency among tasks; waiting to relay a task until another member completes her/his portion of work is due to the dependency among people [10]. One way to mitigate these dependency problems is by supporting dual pathways of coordination through both explicit coordination—purposeful coordination activities through verbal communication—and implicit coordination—unconscious coordination based on shared cognition between team members, such as their awareness of what is known by whom, termed "who-knows-what" in the literature [11, 12]. The role of implicit coordination has gained attention in the last 20 years [13]. Implicit coordination is vital for teams, especially when teams work on fast-paced or volatile projects [12,14], characteristics often found in data science work.

Developing shared cognition about other members' expertise may mitigate coordination problems that originate from members having different expertise. For example, increasing awareness of other members' expertise will develop a member's shared cognition around available expertise in a team and thus empower them to locate expertise when needed (i.e., enabling implicit coordination between members). The term "transactive memory system" (TMS) is used to refer to the shared cognition around others' cognition including their expertise, experiences, and capabilities [15]. One of the key behavioral indicators of teams with a well-developed TMS is role specialization based on a deep understanding and trust of others' cognition [16]. Roles can be used as a means to keep transactive memory intact: when a member joins a new team with the same role they played in a previous team, the new team experiences fewer coordination problems following the addition of the member [17].

In the field of Human-Computer Interaction (HCI), researchers have designed and tested software tools to leverage various coordination mechanisms in a team [9, 10]. For example, in the context of complex analytic tasks such as intelligence analysis, a system that made interim analysis results automatically visible to all team members improved their performance in detecting a suspect [18], likely because of the reduced communication burden. Willett and colleagues showed that imposing a work flow on individuals with no prior data analysis experience could help them generate quality explanations of patterns and anomalies found within a dataset [19]. In this case, by defining an ideal work flow participants had to follow, researchers minimized coordination costs. Yet, no study to date has examined coordination specifically in the context of data science teams, even though data science collaborations have become increasingly common.

Through the findings of this study, I argue that shared team cognition is crucial for the performance of data science teams, especially when those teams are composed of members from multiple functional backgrounds. I also argue that increasing awareness of members' expertise facilitates shared team cognition and hence leverages implicit coordination. I adopted the notion of transactive memory systems (TMSs) as a theoretical lens to account for data science team cognition. The objectives of this research are: (1) To characterize role specialization in data science teams and whether it is based on expertise; (2) To understand the relationship between role specialization and TMSs in real-world data science teams; (3) To understand whether role specialization and/or forming a TMS helps data science teams perform better; and (4) To understand if highlighting inaccuracies in the perceptions of available expertise in data science teams facilitates the development of a TMS. I conducted two studies to achieve these objectives. In Study 1, I researched the relationship between role specialization, TMSs, and team performance by collecting data from participants of a data science competition. I focused on TMSs as a specific type of team cognition that is relevant to data science teams. Study 2 drew on the findings of Study 1 to identify a lever for facilitating TMSs in data science teams.

Organization of this dissertation is as follows. Chapter 2 reviews literature on the practice of data science; selected theories related to role specialization, TMSs and coordination; and the design of technology to aid coordination in work teams. Chapter 3 describes the four research objectives, stated above, and the associated research questions to address them. Chapter 4 introduces the methodology applied in Study 1, an observational study collecting surveys from 36 teams and 74 individuals who participated in a week-long data science competition. Chapter 5 presents the results of the Study 1 based on analyses of the survey data. Chapter 6 details the methodology of Study 2, similar to Study 1, but with an addition of a field experiment focusing on inaccuracies of perceived expertise in teams so as to facilitate the development of a TMS using randomly administered surveys. Chapter 7 presents the results of Study 2. Chapter 8 discusses the implications of the findings from both studies, theoretical and practical contributions of the research, study limitations, and directions for future research.

Chapter 2: Literature Review

2.1 Data Science Collaboration

2.1.1 Definition of data science

Data science has recently risen as a new discipline: it was only in 2015 that the United States federal government officially appointed its first chief data scientist [20]. Data science can be defined as a data analysis practice using statistics [2]. While statistics is a discipline with long-standing history, data science has emerged as a separate discipline from statistics because of the new challenges and opportunities in data analysis introduced by societal changes [2–4]. One of these societal changes includes the explosion of available data on human activities as we leave our behavioral traces in the form of log data every time we connect to the Internet [3]. U.S. governmental bodies have also begun to release data about civic inquiries and administrative actions after the Open Government Initiative was announced in 2009 [21]. Another challenge is the growing portion of unstructured data in the form of text, image, and video data [2]. Novel analytic methods, such as ..., have been proposed and developed in order to analyze such unstructured data. Along with the abundance of the datasets, the open-source software movement has accelerated the democratization of analytic tools. Any interested individuals can acquire and apply analytic techniques packaged as a software library [4].

2.1.2 Data science collaboration

Data science has been described as a team sport [5]. Data science collaboration happens mainly due to demands for various tasks requiring different expertise. For example, a data science unit in an organization often comprises three basic positions—data engineer, data scientist, and data science manager—each in charge of different aspect of data science [22]. In a software company, data scientists take on roles such as insight providers, who inform decisions; modeling specialists, who build predictive models often powered by machine learning models; platform builders, who create data platforms; polymaths, who do all of the above mentioned activities; and team leaders, who supervise teams of data scientists [23]. Civic hackers, a group of volunteer analysts working on open government data, often collaborate cross-functionally by having domain knowledge experts (e.g., government agents, politics or law experts) and technical experts (e.g., software developers) in a same work group [24].

How to build an effective data science team is an open question [22]. One of the topics for discussion is whether to form a data science team with a group of specialists or generalists [6, 7, 25]. In his recent article, Colson argued that the fundamental goal of a data science team is to learn and develop new capabilities, not to "execute" those capabilities [6]. The core difference between the two goals is that former cannot be planned upfront, where the latter may be. When the goal is to learn through multiple iterations, as in data science work, specialization hinders work processes by increasing coordination costs, exacerbating wait time to properly line up subtasks, and narrowing contexts shared between workers [6]. Jurney argued the same, stating that a team of specialists would experience a "Chinese whispers" effect, meaning that streamlining communication across specialists would cause communication overhead [25]. Tunkelang made a distinction between earlier and later stages in setting up a data science team and argued that generalists are more useful for teams at the early stage while, as teams age, the need for specialists may grow [7]. These accounts are based on opinion rather than scientific studies, as there are few scientific studies on data science collaboration.

2.1.3 Software tools to support analytic tasks similar to data science

In the field of HCI, many researchers have developed and tested tools to assist complex analytic work, although most of these tools were not specifically designed to aid data science work. A body of work in visual analytics has devised interactive visualization techniques and platforms with the aim to mitigate cognitive biases of analysts [26], to support parsing geospatial data [27], and to help analysts expand their exploration of variables in their analyses [28].

Only some of these tools were designed specifically to improve collaboration. Automatically sharing an individual's interim analysis, named "implicit sharing" by the researchers, led to participants remembering more clues to solve a given problem than without the sharing function [18]. Balancing the visibility of collaborators' work was found to be key in designing an interface of a tool for collaborative analysis [29]. With structured work flow, a crowd with no prior training in data science could produce high-quality explanations for data analysis results [19].

2.2 Selected Theories of a Team's Role Structure, Cognition, and Coordination

2.2.1 Team coordination & roles

2.2.1.1 Team coordination

Coordination in the context of team work refers to managing dependencies in tasks, resources, and people [10]. Coordination can be either the process of coordination or the state of coordination as an outcome of a process. When coordination breaks in a team, performance is likely to suffer. As the size of a team grows, dependencies among its members will also grow thus coordination in the team becomes more difficult.

Coordination in a team is executed through two pathways: implicit and explicit mechanisms [12]. Explicit coordination mechanisms consist of purposeful actions to coordinate the performance of team members through verbal communications, with or without the help of coordination tools (e.g., calendars, blueprints). Implicit coordination happens when team members unconsciously coordinate based on unspoken assumptions, namely shared team cognition [11]. Team cognition is the cognition of team members about tasks at hand or other members and is grounded in the shared knowledge of team members [12, 30–32].

Although there are numerous ways to coordinate either explicitly or implicitly, an optimal solution for team coordination varies by team and its situation [12]. While maintaining channels of explicit coordination is a key to all teams for smooth collaboration, too much communication may hinder an individual member's productivity due to information overload [33]. Implicit coordination can complement explicit coordination without putting as much cognitive load on members. However, members should have some prior knowledge of both their own and other members' tasks in order to form team cognition, a prerequisite for implicit coordination. As shared team cognition is multidimensional, teams may develop a certain aspect of team cognition more than others. Research in team cognition has conceptualized many different terms to name different aspects of team cognition, though the terms are not used exclusively [34]. Transactive memory system is one of the terms, and it is a group cognition based on knowledge of who-knows-what [15].

2.2.1.2 Role structure for team coordination

Previous studies in social science define a role to be a series of actions which an individual performs [35] and/or to a cluster of behaviors conceptualized for someone in a specific position [36]. Roles may be given to an individual in an organization based on the organizational need or emerge from repeated actions of the individual in a social setting. Either way, once a role is formed and becomes explicit, that role dictates an individual's actions. In other words, an individual with a certain role is expected to perform certain actions aligned with that role. Studies have found that individuals are more likely to act in accordance with given roles when roles become more salient [37].

Roles are the basic constituent for socialization that enable stable social interactions in organizations [38] and role structure can be a driver for team coordination [39]. In a team setting, members often assumes roles that match their personal characteristics or functional expertise [36]. As roles encode individual responsibilities, properly designed role structures can enable coordination between complete strangers and improve performance in existing teams [39, 40]. For example, redesigning role structures in a hospital emergency department resulted in a 40% improvement in patient throughput time [40]. Role-based coordination is especially relevant for fast working teams with open-ended tasks [39, 41].

Certain team roles were claimed to improve team performance according to the Team Role Theory [42, 43]. This theory claims that teams with members who cover nine team roles could achieve higher team performance than teams that do not. The nine team roles include plant, resource investigator, coordinator, shaper, monitor-evaluator, team worker, implementer, completer-finisher, and specialist. Researchers claimed that the theory can be generalized to teams working on various types of problems because the team roles were thought to be distinguished from functional roles based on specialized knowledge [43]. Due to difficulties in measuring team roles and defining team performance in various contexts, this claim has not been tested in diverse enough contexts to support the generalizability of the theory.

2.2.1.3 Roles in data science teams

The debate of hiring specialists or generalists for data science teams [6, 7, 25] can be reframed as a debate of forming a team with members who have specific and specialized roles versus those who have no clear roles. The roles discussed in the cited articles were functional roles determined by one's skill and knowledge [36]. When team members are given clear and specific roles and have a deep understanding of their functions and responsibilities, they can anticipate their interactions with others thus increasing the stability of team coordination [44]. However, when the same debate is argued in by data science practitioners [6, 7, 25], constructing a data science team with specialized roles raises coordination cost due to differences in the use of jargon, interpretation, and motivation [8].

Researchers recently began to study the roles taken by members of data science teams. Harris and colleagues defined 4 roles (i.e., Data business people, data creatives, data developers, and data researchers) of corporate data scientists through surveys on skills and self-identification with various professional categories [45]. Kim and colleagues took on a systematic approach and defined 9 roles by collecting time spent data from employees working on data science fields in Microsoft and performing a clustering analysis with the time spent data [46]. They identified 9 clusters that are polymaths as generalists who would engage in all activities, data evangelists who actively support data-driven decision making in the organization, data preparer who spent their time mostly querying for existing data, data shaper who mostly prepared data, data analyzer who mainly analyzed a given data set, platform builder who spent time in building platforms for data management, fifty-percent and twenty-percent moonlighters who spent their time doing activities irrelevant to data science, and insight actors who spent their time in executing actionable insights from data analysis. Saltz and Grady assembled case studies to identify a set of roles related to data science including data science researcher, data scientist, data architect, data analyst, data science programmer, and data engineer [47]. Zhang and colleagues discussed how employees at Microsoft collaborated to tackle data science problems around give major roles (i.e., engineer/analyst/programmer, researcher/scientist, domain expert, manager/executive, and communicator) though no description on how they came up with the roles has been provided [48].

2.2.2 Team Cognition: Transactive Memory System

2.2.2.1 Definition of transactive memory system

A transactive memory system (TMS) refers to a mechanism through which a group of individuals encode, store, and retrieve information and knowledge [49]. A TMS draws on the idea that a group mind works as organically as the mental operations of an individual. TMSs have been applied to describe the shared cognition of team members' knowledge of each other, one sub-dimension of team cognition [34]. According to the research on TMSs, in a group of people who maintained close relationships for a long time (e.g., a family, a romantic couple, or a work team in an organization) the members eventually develop meta-knowledge of who-knows-what over time. For example, in the context of a dyad working in a data science project, member A has experience in analyzing civic data and member B has expertise in building interactive dashboards. As the team continues to collaborate, A and B are likely to divide tasks based on their knowledge of each other's expertise.

2.2.2.2 Antecedents of transactive memory systems

TMSs have been extensively studied in the context of work teams since the term was first coined in 1985 [15]. Ren et al. argued that a TMS for a team within organizational boundaries has three layers of antecedents: team member attributes, such as demographics or technical competence of members; team-level inputs, such as task/goal interdependence or familiarity among team members; and organizational-level inputs, such as geographic dispersion of members [50].

Personal attributes like gender, ethnicity, and certifications as signals of experience affect how members perceive the expertise level of others, thereby affecting the formation and application of a TMS [51]. In the study by Hollingshead and colleagues, participants relied on gender stereotypes to speculate on the relative knowledge of other collaborators. These assumptions affected the division of subtasks in a subsequent task [51]. In multidisciplinary teams within which members are professionals in distinctive areas, members experience challenges originating from differences in the use of language, interpretation, and other discipline-specific conventions such as task procedures [8, 52]. These challenges negatively relate to the formation of TMSs. However, guided interactions help to overcome the challenges in multidisciplinary healthcare research teams [52]. The more an individual identifies with their professional group, the higher a TMS was developed, even when they identified less with their teams [53].

Multiple researchers have argued that task interdependence, an attribute of team-level inputs whereby the degree to which individuals' work outcomes are interrelated to each other's, is a major influence on the development of a TMS [15, 54]. Perceived goal interdependence as well as perceived task interdependence is positively associated with a higher level of a TMS, leading to improved team performance [55]. Group, rather than individual, training is another antecedent that advanced TMSs in multiple studies [56, 57].

Developing a TMS can be more beneficial for certain types of tasks than others; researchers have conceptualized this as "TMS relevance," meaning that the relevance of a TMS is variable by task type [58]. Research on which task types benefit more from higher TMSs, however, is contradictory. In their review paper, Lewis and Herndon found that the strength of the TMS-performance relationship is strong in production and assembly tasks, which have a modular task structure and low interdependence between subtasks [58]. Conversely, Akgun and colleagues found that the TMS-performance relationship was greater for more complex tasks, defined as being less repetitive and requiring more novel knowledge [59].

Research on the relationship between role structure and TMSs in work teams is scarce yet a few studies consistently point towards the negative association between the two constructs. In a team of experts from different fields, the different use of jargon and varying motivations were negatively associated with the development of a TMS in the team [52]. Teams with compartmentalized knowledge developed less converged TMSs than teams with homogeneous knowledge [60]. Teams with members in charge of similar functions developed a higher level of TMSs than teams with members holding different expertise when asked to work on computer-simulated decision-making tasks [61]. Roles can be used as a transmitter of a TMS between teams. A team of strangers were able to develop a TMS when the role structure from the previous collaboration persisted in the new team [17]. With the same role structure, highlighting the role division during team communication was associated with improved team performance with the mediation effect of a TMS [62].

Of the numerous antecedents of TMS development, researchers are increasingly paying attention to the role of information technology (IT) in developing TMSs in work groups [58]. Human memory is known to adapt to the advent of new technology that makes information readily available at our fingertips [63]. As IT is increasingly used to support work groups, such as databases and search engines, these technology systems can substitute TMSs by making TMS processes (i.e., encoding, storing, and retrieving knowledge) more efficient [50, 58]. For example, an organization's document repository can be used as an expertise directory to source expertise on demand as well as an internal knowledge repository [64, 65]. Perceived level of IT support in an organization positively affects the development of TMSs in its teams [64]. The exact mechanism of how IT should be designed to mediate or facilitate TMS development is yet to be explored however.

While IT may facilitate TMS development in teams where members are col-located, multiple sources of evidence suggest that virtual teams (i.e., members are geographically distributed and can only connect through IT support) often struggle to develop an effective TMS due to the lack of nonverbal cues or missed opportunities for shared experiences [66–68].

Some studies have held that TMSs can be artificially created for groups by assigning members to a specific task [62, 69, 70] or by providing information about the expertise of other members [57, 71]. Imposing predefined task structures to impromptu couples—a pair of strangers—is effective in increasing their recall accuracy in a collaborative memorizing task, though it hinders the performance of real couples [69]. Compared to when teams were forewarned that they had more information about one suspect than other team members, one study found that revealing which team member had additional information to all members improved the accuracy of the group when identifying the suspect in a mystery crime scenario [70]. Providing expertise information is especially helpful to TMS retrieval by scaffolding expert identification process. For example, in a radio assembly task, a TMS was artificially created by revealing other members' expertise; those in the TMS team outperformed those who had the benefit of improved team communication [57]. When allocating tasks to match the requisite expertise, team member's expertise information should be considered for accurate matching [71]. In terms of how to provide expertise information, Brandon and Hollingshead proposed a "task-expertise-person" (TEP) unit as a mental constituent of a TMS, although they did not examine if providing TEP information to team members would improve their TMS [54]. These findings on the effectiveness of artificially-created TMSs suggest that IT can enable TMSs.

2.2.2.3 Benefits of transactive memory systems

Within teams with well-developed TMSs, members have a reduced burden of memorizing new information relevant to a given task in the belief that other members responsible for the information will also hold the memory [49]. As a result, this reduced cognitive load and efficient division of responsibilities lead to the enhancement of team performance [16, 50, 57, 72, 73]. Large size teams benefit from TMSs especially when they are in a dynamic knowledge environment where task and requisite knowledge often alters during execution [74].

Some researchers have argued that TMSs also provide other benefits to teams [17,75–77]. Top management teams with well-developed TMSs can pursue ambidextrous orientation, meaning that they explore and exploit necessary information more flexibly [75]. A TMS enables smooth coordination among strangers when members are assigned to the same roles that they had assigned in the previous collaboration [17]. In this situation, a TMS was encoded in roles and could thereby be transmitted to a new team. Additionally, teams can withstand a loss of an existing member better when the lost member had played a role insignificant to team work [77].

2.2.2.4 Transactive memory system in Data Science Teams

No research to date has studied TMSs in data science teams, though one may infer from previous studies on TMSs in teams working on tasks similar to data science. A simulation study showed that a TMS enabled teams to cope with volatility by allowing members to proactively adjust their behaviors based on the team cognition [74]. At the same time, data science work requires a breadth of knowledge from various fields such as statistics, programming, or social sciences; mastering all of these fields is a strenuous job for an individual. As members of a data science team are expected to add unique knowledge to a team, it may become harder for a team to develop a TMS due to knowledge boundaries, such as differences in the use of jargon and interpretation [52].

2.3 Software tools to support team coordination

HCI research has a number of studies on designing software tools to support human collaboration [78,79]. Coordination Theory was coined from the perspectives of HCI, with an aim to develop a framework to understand team coordination and to design a software tool supporting that coordination [9]. In their seminal paper, Malone and Crowston defined coordination as management of dependencies among tasks, resources, or people [9, 10]. Examples of such dependencies are sharing the same hardware tools with limited resources, waiting to relay a task until another member completes their portion of work (i.e., producer/consumer relationship), and waiting to edit a draft until some content is filled (i.e., task/subtask relation) [10].

Some of these dependencies can be managed more effectively using software tools. If you have to wait to work on task B until another member is done with task A, a simple notification software programmed to ring an alarm when the member finishes task A would give you time to work on other tasks without being concerned about when task A will be done. If team size grows and dependencies multiply, having a shared dashboard that displays the progress of each member and each task could significantly reduce the information overload a member may experience as the team size grows [80].

Such software tools can support either explicit or implicit coordination, or both simultaneously. For example, when a group of voluntary editors edited an existing article or created an initial draft on Wikipedia[1], the largest online encyclopedia created from the work of volunteers, they coordinated their activities explicitly on an online forum or using a text-based messenger. At the same time, they coordinated implicitly by building on an initial content (i.e., skeleton draft) instantiated by a few committed editors [81].

One line of research designed and tested software tools for complex analytic tasks such as intelligence analysis. Automatically sharing an interim analysis result to another team member improved the performance of a pair working to detect a suspect [18]. Some studies built on an existing software tool commonly adopted for data science, namely "Jupyter Notebook"[2] [82]. When enabling users to fold (hide) a chunk of multiple cells and to leave a header as an annotation for the folded chunk, a group of users navigated a relatively long Jupyter Notebook script more easily than when the feature was not enabled [83].

[1] http://wikipedia.org

[2] Jupyter Notebook is an interface that allows users to run several scripts on a single page thereby to set up an analytic work flow easily

Chapter 3: Research Objectives & Contribution

3.1 Research Objectives

This dissertation has four objectives: (1) To characterize role specialization in data science teams and whether it is based on expertise; (2) To understand the relationship between role specialization and TMSs in real-world data science teams; (3) To understand whether role specialization and/or forming a TMS helps data science teams perform better; and (4) To understand if highlighting inaccuracies in the perceptions of available expertise in data science teams facilitates the development of a TMS.

The research model of the dissertation is shown in Figure 3.1. Each link between the constructs notated with a corresponding research question. Each research question is described in the following section.

To achieve the research objectives, I conducted two studies. Throughout the two studies, I applied TMS theory as a lens to account for team cognition, a potential basis for implicit coordination in data science teams. Study 1 was an observational study which investigated 36 student teams of two to four members who competed in the week-long University of Maryland 2019 Data Challenge. Through surveys, I collected information about students' expertise, task division within teams, and

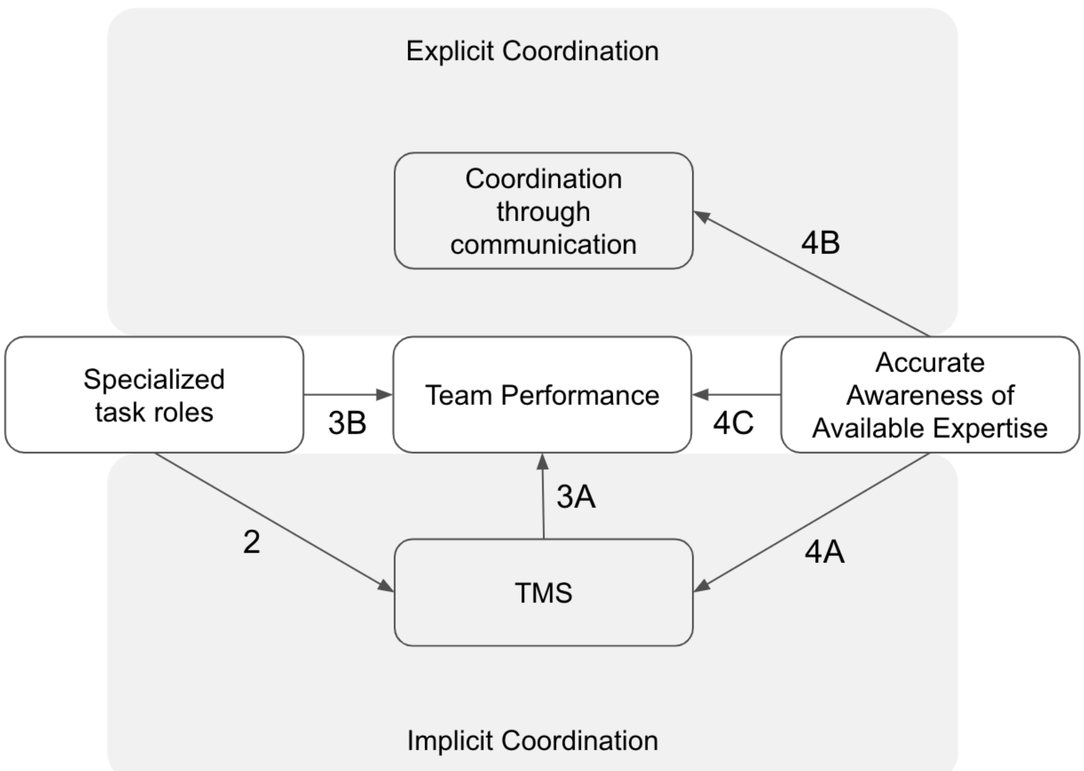

Figure 3.1: The research model of the dissertation

whether they reported forming a TMS. In addition, I gathered team scores from the competition. The goal of Study 1 was to understand whether and how teams divided up tasks into roles, whether teams developed a TMS, and whether these two aspects of teamwork predict performance in the competition.

Study 2 was an experimental study which tested if highlighting inaccuracies in perceived expertise to team members would help them develop a TMS. The expectation was that highlighting inaccuracies would incur synchronization on the available expertise among team members. Having all team members on the same page was expected to develop a TMS and, in turn, to improve team coordination and performance.

Research Objective 1: To characterize role specialization in data science teams and whether it is based on expertise

Teamwork is an integral part of the practice of data science [5,22]. I argue that role specialization of members in a data science team plays an important factor for coordination and performance of the team. Data science teams need to pool knowledge from various fields such as software engineering, statistics, and social sciences and it is difficult for any one individual to master all of these areas of knowledge [22]. Members in these teams are expected to bring unique knowledge to the team and complement or expand the existing knowledge pool. This composition of experts from different fields may lead to divided responsibilities of labor [61]. Despite the potential benefits of this approach, data science practitioners have defended teams comprised of generalists due to concerns over increased coordination cost between experts from distinctive fields in teams with specialists [6, 7, 25]. Yet, no empirical evidence has shown how data scientists divide tasks and specialize in certain task roles, if they do at all. In this research, I collect empirical evidence to understand the role behavior in data science teams. The goal is to understand whether members in data science teams specialize in roles based on their relative expertise. I address two research questions as follows.

Research Question 1A: Are tasks divided among members of data science teams to create differentiated roles within teams? If so, how? Research Question 1B: Is team member expertise related to which tasks someone completes?

Study 1 and Study 2 addressed this objective by identifying roles (if any) based on task division and testing if there was an association between roles and expertise.

Research Objective 2: To understand the relationship between role specialization and TMSs in data science teams

I argue that role specialization negatively predicts TMS formation. Role specialization in refers to the division of functional roles observed from actual task delegation in teams; TMSs refer to each member's cognitive perception about whether other members specialize in certain roles. Results from previous studies on the relationship between teams with different functional roles and the development of TMSs have been consistent: different use of jargon and varying motivations are negatively associated with TMSs in teams of experts from different fields [52]. Teams with members from similar fields develop a higher level TMSs than teams with members with different areas of expertise in computer-simulated decision-making tasks [61]. Teams given differentiated knowledge for a task end up developing a less converged TMS than teams given similar knowledge [60]. No study has examined the relationship between role specialization and TMSs in data science teams. I address the following research question to fulfill this research objective.

Research Question 2: Does role specialization help or hurt data science teams to develop transactive memory systems?

Study 1 addressed this objective by analyzing the relationship between role specialization in teams and the development of TMSs in teams participating in a week-long data science competition. Study 2 addressed this objective by analyzing the same relationship between the same constructs, but in an experimental setup.

Research Objective 3: To understand whether role specialization or forming a transactive memory system helps data science teams perform better

I argue that role specialization itself has no effect on the performance of data science teams unless teams have prior experience of working together. A team of experts from distinctive fields may be able to pool a breadth of knowledge from its members when members have a certain level of understanding of each other's expertise [60]. However, a team of strangers would lack such knowledge and thus may not benefit from having experts from multiple fields.

Drawing on evidence from prior research, I argue that TMSs would improve performance in data science teams. Studies have showed that teams with well-functioning TMSs were more likely to pool available resources and expertise and hence perform better than teams with poorly developed TMSs [55, 56, 84, 85].

Again, no work to date has collected empirical data or tested the relationship between role specialization and performance or between TMSs and performance in data science teams. I examine these relationships by addressing following research questions.

Research Question 3A: Do TMSs improve performance in data science teams?

Research Question 3B: Does role specialization improve performance in data science teams?

Study 1 addressed this objective by testing the association between a TMS, as measured by self-reports, and team performance in the data challenge. Study 2 developed and tested an intervention to improve TMSs in data science teams. Study 2 experimentally tested whether teams given an intervention also improved their performance and whether that was mediated by the development or strengthening of a TMS.

Research Objective 4: To understand if highlighting inaccuracies in perceptions of team member expertise facilitates the development of transactive memory system, the amount of explicit coordination, and team performance in data science teams

I argue that aligning expectations about the available expertise in teams would be beneficial to data science teams. When information that could be useful for team work is not distributed between members of a team, the team loses a chance to improve the outcome of their collaboration [86]. The expectations about available expertise could be aligned by first recognizing any inaccuracies that may exist in members' perceptions of other members' expertise. I hypothesize that by highlighting any inaccuracies in perceptions of team member expertise in data science teams, the members would be able to recognize the knowledge resources available within their teams and to develop a TMS faster.

Research Question 4A: Does highlighting inaccuracies in perceptions of team *member expertise improve the development of transactive memory systems in data science teams? Research Question 4B: Does highlighting inaccuracies in perceptions of team member expertise increase the amount of explicit coordination in data science teams? Research Question 4C: Does highlighting inaccuracies in perceptions of team member expertise improve team performance in data science teams?*

Study 1 did not directly address this objective. Instead, the list of daily tasks participants reported in Study 1 informed the design of Study 2 by offering an overview of how teams divided work and what task roles individuals assumed. Study 2 addressed this objective through a field experiment that highlighted any inaccuracies in perceived expertise among participating team members.

Chapter 4: Study 1 Methodology

4.1 Research Methods

The UMD Data Challenge is a week long competition hosted annually by the College of Information at the University of Maryland (UMD), College Park. I collected data from the second Data Challenge event which was hosted between February 23rd and March 2nd, 2019. In the competition, undergraduate and graduate students from UMD competed to create the most innovative and technically sound data science project. Students could participate as an individual or in a self-formed team of up to four members. Teams and individuals were given a range of datasets to choose from, worked with the datasets for one week, and then presented and submitted a final analysis of the data. Final presentation and submissions were evaluated by external judges who assessed the relative strengths and weaknesses of the data science projects and selected a winner.

4.1.1 Participants

Participants were recruited among the students who elected to participate in the UMD Data Challenge 2019. Out of 61 teams and 188 students who participated the competition, 133 students from 53 unique teams consented to be a part of this study. By the end of Data Challenge 2019, I collected data from 39 unique teams and 81 individuals. Of the 81 individuals, 77 completed and 4 partially answered the final survey. Consequently, in my analysis, I included 36 teams and 74 individuals from teams with at least two members (it was not required that all team members consent if an individual wished to participate in the study). All student participants were rewarded with a $10 Amazon gift certificate for completing the surveys.

Of the 74 participants, 44 students were female and 30 were male. Thirteen were undergraduates, 58 were graduate students in masters programs, and 3 were Ph.D. students. The top five of most common majors were Information Science, Business Analytics, Information Systems, Marketing Analytics, and Computer Science and Engineering.

4.1.2 Study Design

Study 1 was an observational study. Student participants were free to form a team as they wished as long as the size was not over four. I measured individuals' areas of technical expertise prior to the actual competition period (February 23rd to March 1st, 2019). I measured the state of each team's TMS by gathering data on team cognition and task delegation at the end of the competition when student participants gathered for final judging (March 2nd, 2019).

4.1.3 Materials

I collected data through multiple surveys, both before and during the Data Challenge competition. These surveys were: an intake survey, daily diary surveys, and a final assessment survey. Judges' ratings data was also collected and contained evaluation and results of final project outcomes.

Daily diary surveys asked each participant what tasks they did ("What were the activities you did today related to the Data Challenge? List as many activities as possible each as a separate entry below in the order that you worked on them."), how long they spent on the task ("Approximately how many minutes did you spend on this activity?"), and who they worked with, if anyone, on the task ("Did you work on this activity alone or with others (include your mentor and anyone you were actively discussing the activity with at the time)?").

The final survey asked student participants about the degree to which they developed collective memory on which team member knows what (i.e., their TMS). This survey was previously used and validated by Lewis (2003) to measure behavioral indicators of TMSs. The final survey also asked about the experience of working in a team such as how participants formed their team ("Which of the following best describes how well you knew your team members before participating in the Data Challenge?"), task division among members ("For each of the following types of activities write down which team members worked on this part of the project and what days it was worked on."), team interdependence [87], and team satisfaction [88].

Administrative data included profiles of participants collected from Data Challenge registration forms and judges' ratings on final projects presented on the last day of Data Challenge 2019. When student participants registered to participate in the Data Challenge, they were asked to provide demographic information such as years of experience/education in a major and/or skills related to data science (e.g., statistics, visualization, design, Python libraries for data science, R). This information was used to measure the diversity of knowledge backgrounds within teams. At the end of the competition, three judges per team evaluated all final submissions made by participating teams with 22 total judges. All judges were given a rubric as well as instructions for the evaluation. The rubric measured five dimensions of quality: completeness of the project, quality of analysis result, innovativeness of the project, quality of presentation, and consideration for community integration, the theme of Data Challenge 2019. Each dimension of the rubric had four questions, each of which asked judges to give a rating between 0 and 10.

4.1.4 Procedure

Participants were asked to provide their demographic information along with their degree major, degree program, skills, and motivation for participation in the intake survey shown during the registration process for the Data Challenge competition. They were also asked whether they would participate in this study by completing consent forms.

Daily diary surveys were administered to collect the time-use information of participants everyday during the competition. Student participants who consented to be contacted by email and/or text were sent a link to an online daily diary questionnaire hosted by Qualtrics at 9pm each day of Data Challenge 2019. When students forgot to submit a daily diary survey for a given day, I sent reminder emails or texts to the students at 9am the next day.

For the final survey, student participants were contacted one last time during the judging day before results were announced. They were asked to fill out the final survey and to complete any missing daily diaries.

After completing the final survey, participants were given a debriefing message.

4.1.5 Data Analysis

In this section, I address, in turn, data analysis methods used to answer each research question.

4.1.5.1 Research Question 1A: Are tasks divided among members of data science teams to create differentiated roles within teams? If so, how?

To answer research question 1A, I analyzed how teams divided task roles, if any, by performing K-means clustering analysis on the data about how teams delegated five sub-tasks of data science work (formulating a research question, cleaning data, exploring data, modeling data, and presenting results). Student participants were asked to provide a list of tasks they worked on for Data Challenge 2019 for each day during the competition. Participants also reported how long they worked on each task and with whom they worked (by providing initials of the collaborators).

The collected data were normalized and imputed for missing values prior to the actual analysis. First, I normalized the inputs on day of the week names (e.g., "all the days" to each day during the competition period) and group member initials. The competition spanned two Saturdays: I differentiated the work reported to be completed on the second Saturday from the work done on the first Saturday by assuming individuals report the days of the week in order (e.g., when the input was "Friday, Saturday," it was assumed to be the second Saturday). Second, missing entries on time information in daily diary surveys were imputed by referring to automatically generated time-markers at the moment of each survey submission.

Individual level responses were aggregated at a team-level following several steps. First, two researchers, including the author, merged entries on the same task reported by different members to generate a list of all tasks reported by the team. If two or more entries had similar descriptions (e.g., "Brainstorm the questions", "The team brained stormed all kinds of questions"), were reported by more than one member (i.e., collaborative), and reported to be done on the same day, we deemed that the entries referenced the same task. After generating a complete list of all tasks done by every team, two independent raters, including the author, coded each task submitted in free-text format. Codes for the tasks were determined by referring to the previous literature on the taxonomy of data science work [] and by open coding and discussion between the two raters. The final task codes are: 1) Orienting and developing research questions; 2) Wrangling and cleaning data; 3) Exploring data; 4) Analyzing data; 5) Interpreting results; and 6) Other tasks.

Definitions and details on the codes are shown in Table 4.1.5.1.

Task Code	Definition
Developing research questions	Helped generate research question(s) or goal(s)
Wrangling and cleaning data	Cleaned, manipulated, and/or validated the data
Exploring data	Explored, summarized and/or visualized the data
Analyzing data	Built statistical and/or machine learning models using the data
Interpreting results	Interpreted and/or summarized the results
Other tasks	Any tasks that do not fit in another task category

After independently coding each task, the raters consolidated the codes. All disagreements in coding were settled into a final version after discussion between the raters. The final dataset contained the tasks 188 student participants completed for Data Challenge 2019.

After cleaning quantitative data, clustering analysis based on K-means algorithms was performed to identify if task delegation patterns could be grouped into clusters. Each cluster would consist of a set of tasks frequently performed together, thus it is natural to conclude that each cluster corresponds to a functional role. I chose the K-means clustering algorithm because K-means clustering is robust against outliers and irrelevant variables [89, 90]. The Davies-Bouldin clustering index was used to determine the optimal number of clusters. This index is a widely-used indicator to determine the number of clusters in cluster analysis [91]. The rule of thumb is to pick the number of clusters with the lowest Davies-Bouldin index value because a lower Davies-Bouldin index represents better distinction between clusters (i.e., inter-cluster) and more cohesiveness within each cluster (i.e., intra-cluster). First I normalized the data at individual level to adjust for individual differences in commitment to the competition. Without normalization, the clustering results would be biased towards highly committed participants. Then I clustered the data using the K-means algorithm into a minimum of two and a maximum of 10 clusters. Ten clusters were regarded as the maximum, since it is the maximum number of possible combinations when a role consists of two sub-tasks out of five possible sub-tasks.

4.1.5.2 Research Question 1B: Is team member expertise related to which tasks someone completes?

Independent Variable

Relative competency in areas related to data science: An individual's relative competency was measured in several areas of expertise related to data science including subject matter, programming, statistics/mathematics, and data science project experience. Relative competency was operationally defined as their competence compared to the other team members in a given domain area. For example, member A may have 5 years of experience in public policy, and the sum of years of experiences in A's team is 8 years. In this case, A's relative competency was calculated as 0.63 [1].

Dependent variable

Specialized role: A specialized role was operationally defined as a set of tasks that share similar characteristics. I drew on the result of the clustering analysis in

[1] An individual's years of experience over the sum of all team members' years of experience in a given domain

the research question 1A to determine if a member played a specialized role. In the clustering analysis, the algorithm identifies a cluster for each unit (i.e., individual participant) that has the shortest euclidean distance between the data pattern and the centroid of the pertinent cluster. To decide if a role was specialized in a team, I followed these decision criteria: (1) If half or less than half of team members (i.e., one or two members at most depending on the team size) were in charge of a group of tasks identified as a role, the member(s) was considered to be specialized in that role and (2) if only one member was in charge of a task, that person was considered to be specialized in the task. Guideline (2) is to account for a role structure specific to a team because not all teams would share the same role structure.

To answer research question 1B, I hypothesized that expertise, especially external expertise, would be associated with the assumed roles. For example, if a member was more competent in developing machine learning models compared to other team members, that member would likely to take on a role of analyst. To test this hypothesis, I fitted the data to a multinomial logistic regression model to see if having relative expertise in a certain area is associated with taking a specialized role that would require relevant experience and expertise to execute it.

4.1.5.3 Research Question 2: Does role specialization help data science teams to develop transactive memory systems?

Independent Variable

The number of unique roles in team: To capture how role specialization manifests at team level, I aggregated the number of unique roles.

Dependent variable

A Transactive Memory System: I adopted questionnaire developed by Lewis [16] to measure the TMS in each team. The questionnaire had a total of 15 questions, five for each of the three behavioral indicators of a TMS: having trust in other members' work, experiencing less frictions in coordination, and specializing in a certain role.

Control Variable

Team size: I included the number of members for each team in the analysis to parse "size effect" from the effect of role specialization. Size effect means that merely having more members to help out in a project may lead to higher team performance. In this study I hypothesized that having members from diverse backgrounds to pool knowledge and skills from may lead to higher team performance, name a team diversity effect.

Statistical test

I hypothesized that having specialists in teams would hinder the development of a TMS. To examine this hypothesis, I fitted the data using a multiple linear regression model.

4.1.5.4 Research Question 3A: Does transactive memory system improve performance in data science teams?

Independent Variable

Transactive memory system: This is the same variable measured for the dependent variable in the research question 2.

Dependent variable

Team performance: Team performance was measured as the average score given by judges across five sub-dimensions of performance. The five sub-dimensions were completeness of the project, quality of analysis result, innovativeness of the project, quality of presentation, and consideration for community integration, a theme of Data Challenge 2019. Team performance in each dimension was measured with four questions on the scale of 0 to 10. I averaged the ratings across the four questions and across the three judges for a sub-dimension of team performance.

Control Variables

I included the control variables defined for research question 2 to control for factors that potentially confound the relationship between a pre-existing TMS and team performance.

Statistical test

I hypothesized that the extent of TMSs developed in teams would be positively associated with team performance. The hypothesis was tested using a multiple linear regression model.

4.1.5.5 Research Question 3B: Does role specialization improve performance in data science teams?

Independent variable

Divisional role structure with specialized roles: This is the same variable measured for the independent variable of research question 2.

Dependent variable

Team performance: This variable is the same as the dependent variable of research question 3A.

Control variables

I included the same set of control variables defined for research question 2, in order to control for factors that potentially confound the relationship between role specialization in teams and team performance.

Statistical test

I hypothesized that having specialists in teams would be positively associated with team performance. The hypothesis was tested using a multiple linear regression model.

Chapter 5: Study 1 Results

In Study 1, I addressed three research questions divided into five sub-questions to achieve three research objectives. The overall aim of the study was to observe data science collaboration with a focus on role assignment—in other words, division of responsibility—based on expertise and its impact on the development of a TMS and team performance. The aim was driven by the paucity of literature describing the characteristics of data science collaboration at work.

5.1 Research Question 1A

Are tasks divided among members of data science teams to create differentiated roles within teams? If so, how?

Individuals in a group often divide their responsibilities for labor, in other words task roles, for collaboration [15,49]. Dividing task roles in a team has multiple benefits for team work. By dividing roles, teams can shorten the time spent to complete a task and/or expand knowledge resources to draw from when tackling a complex problem [42]. I expected that data science teams would be composed of members with varying task roles. Knowledge in multiple domain areas is often required to tackle data science problems [5,7]. To see if this expectation is in line with empirical evidence, I conducted a K-means clustering analysis based on reported task division in teams to quantitatively describe the types of roles individuals were adopting in teams. In order to determine the optimal number of clusters, I referred to the Davies-Bouldin index, a commonly used metric to evaluate the quality of clustering analysis results [91, 92]. The score was the lowest at 4 clusters with the value of 0.82 (See Table 5.1).

Cluster Size (K)	Davies-Bouldin Score
2	1.03
3	1.03
4	0.82
5	0.84
6	0.85
7	0.92
8	0.85
9	0.86
10	0.90

Table 5.1: Results from K-means cluster analysis on task division. Davies-Bouldin Score per number of clusters ranging between 2 to 10

Each cluster has a centroid, a vector of the means of all variables for observations in a given cluster. A centroid can be thought to describe an archetype of a cluster [93]. Based on the values of each centroid, I qualitatively defined four functional roles to which each cluster corresponds: generalist, research question generator, wrangler, and analyst (See Table 5.1 to match each role to a corresponding centroid). On average, individuals clustered into the *research question generator* role spent most of their time developing research questions by doing tasks like background research or generating initial research questions and refining them (0.75). They spent the remaining time cleaning data (0.15). *Wranglers* performed data

Cluster name	Developing RQ	Cleaning data	Exploring data	Analyzing data	Presenting findings	Other tasks	% of individuals
Generalist	0.14	0.2	0.16	0.1	0.2	0.2	69.5%
RQ generator	0.75	0.15	0.0	0.01	0.01	0.08	12%
Wrangler	0.03	0.85	0.11	0.0	0.0	0.01	9%
Analyst	0.06	0.09	0.02	0.74	0.1	0.0	9%

Table 5.2: Centroid values of the four clusters representing the four task roles recurring in Data Challenge 2019

cleaning tasks like reformatting, sub-setting, or combining data (0.85), followed by data exploration tasks like getting descriptive statistics or visualizing the data (0.11). Individuals categorized into the *analyst* role used the majority of their time doing inferential, spatial, and/or text analysis and/or building machine learning models (0.74). A fraction of their time was spent cleaning data (0.09). The *generalist* role, the group with the most individuals, engaged evenly across all tasks with centroid values ranging from 0.14 to 0.2.

Over two-thirds of participants (82/118; 69.5%) were engaged across all tasks, thus were categorized as generalists according to the results of the clustering analysis. Only 14 (12%) participants were clustered as research question generators. Wranglers and analysts were 11 (9%) and 11 (9%) respectively.

Out of a total of 30 teams, eight (26.7%) had differentiated roles within the team by having two or more members with different roles. This finding suggests that individuals may have specialized in different sets of tasks; however, most of these differences reflected between-team differences in the types of tasks individual members performed. Of those eight teams, the most commonly found combination of roles was the generalist and the research question generator (in five teams, 16.7%). Of the other three teams, two had the combination of generalist and wrangler and one of analyst, research question generator, and wrangler.

5.2 Research Question 1B

Is team member expertise related to which tasks someone completes?

Task roles may require prerequisite knowledge to be performed [94, 95]. For example, if an individual is to take on an analyst role, the individual must possess expertise in theories and techniques about data analysis. As data science work is a combination of individually specialized sub-fields such as statistics or programming, I projected that strong ties between the members' expertise and task roles would exist in data science teams. To test the projection, I conducted multinomial logistic regressions analysis with the roles identified from the previous research question (i.e., research question 1A) and the self-reported expertise in order to understand how years of experience in different skills related to the set of tasks an individual performed.

The years of experience of participants across domain areas ranged from 0 to 10. On average, participants had the most experience in getting descriptive statistics (mean: 2.02 years, std: 1.91). Participants had the least experience in front-end programming (mean: 0.7 year, std: 1.19) as shown in Table 5.2.

	Descriptive	Visualization	Data Wrangling	Inferential statistics	Predictive statistics	Database management	Front-end programming	Domain knowledge
Mean (Year)	2.02	1.65	0.87	1.33	0.73	1.03	0.7	0.86
STD.	1.91	1.56	1.2	1.28	0.83	1.08	1.19	1.2

Table 5.3: Means and standard deviations of participants' experience level in eight domain areas related to data science

An individual's expertise often works as a basis for what role the individual would take on in a group or an organization, for example when hiring for a software engineer position in a team at an IT company. I hypothesized that a strong tie would be observed between expertise and role in data science teams. For example, an individual who have many years of experience in inferential statistics would likely take on an analyst role in a team. The hypothesis was examined using mixed-effects logistic regression models. The independent variables were expertise measured in years of experience in 7 domain areas related to data science with four types of roles as a dependent variable. To account for the individual's relative competency over the other team members, I centered all independent variables to the group mean in all models. Three logistic regression models were defined and fitted to the data, each model comparing the odds ratio between two roles given a combination of expertise level (See Table 5.4).

Table 5.4 shows the association between relative competency and task roles. All expertise variables were centered at the mean value of a given team to account for relative competency of an individual over the average in the team. As in the previous model, relative competency failed to significantly account for the odds of taking on a role over the other role.

In summary, the multinomial logistic regression analysis revealed that an individual's expertise was not related to which sets of tasks the individual completed during data science collaboration. Several possible explanations for these results may exist: One, roles other than the generalist were too few so the power to detect any signals was too small. Two, most participants were students thereby had relative limited experiences to take on any specialized roles based on expertise. Further explanations are discussed in Chapter 8.

	Dependent variable:		
	Generalist as reference		
	Analyst	RQ generator	Wrangler
Intercept	-16.149^{*}	-4.138^{***}	-13.390
	(9.678)	(1.220)	(9.104)
YOE: Descriptive stats	-2.263	-0.012	-1.190
	(4.467)	(0.783)	(1.645)
YOE: Visualization	0.113	-0.551	3.307
	(2.435)	(0.971)	(2.011)
YOE: Front-end programming	1.672	0.968	-0.535
	(1.863)	(0.696)	(1.889)
YOE: Inferential statistics	-0.290	-0.615	4.150
	(7.881)	(1.148)	(6.212)
YOE: Predictive modeling	8.221	-1.840	-12.533
	(11.642)	(2.190)	(16.538)
YOE: Database management	-0.642	-0.062	-0.488
	(2.274)	(0.803)	(1.987)
YOE: Data wrangling	-4.438	0.057	4.821
	(8.965)	(1.154)	(6.255)
Observations	73	67	69
Log Likelihood	-8.813	-11.254	-6.808
Akaike Inf. Crit.	35.626	40.508	31.616
Bayesian Inf. Crit.	56.240	60.350	51.723

Sample size (N) = 83

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5.4: Coefficients and standard errors from a multinomial logistic regression with relative competency in data-science related skills compared to team members as independent variables and task roles identified from the clustering analysis results as a dependent variable

5.3 Research Question 2

Does role specialization help data science teams to develop transactive memory systems?

A member specialized in a domain area contributes to the team by sharing their expertise. With multiple specialized members, a team is expected to be able to pool from a broader breadth of knowledge than a team with members taking on generalized and overlapping roles. I expected that role specialization would be associated with the development of TMSs in teams because members would be more likely to recognize if other members were specialized in certain roles, since recognizing how responsibilities of labour is divided in teams is one of the main manifestation of a TMS [74]. To test this expectation, I conducted multiple linear model analysis with the number of unique roles in teams as an independent variable and average TMS score measured on the last day of team collaboration as a dependent variable. According to the result, having more diverse roles in team did not lead to a higher level of a TMS as shown in Table 5.5.

5.4 Research Question 3A

Does transactive memory system improve performance in data science teams?

The development of TMSs in teams has been commonly found to be associated with team performance in various settings [64,96,97]. In order to test if this holds for data science teams, I conducted a linear regression analysis with the average TMS

	Dependent variable:
	Development of TMS in team
Intercept	3.002***
	(0.611)
Number of specialized roles in team	0.027
	(0.116)
Team size	0.165
	(0.163)
Observations	30
R^2	0.042
Adjusted R^2	−0.029
Residual Std. Error	0.372 (df = 27)
F Statistic	0.596 (df = 2; 27)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.5: Coefficients and standard errors from linear regression analysis with the number of unique roles in team as an independent variable and average TMS score measured on the last day of collaboration as a dependent variable

score as an independent variable and team performance as a dependent variable. The result shows that the extent to which TMSs developed in teams was not statistically associated with changes in team performance as shown in Table 5.6.

	Dependent variable:
	Team performance
Intercept	43.071***
	(8.946)
Development of TMS in team	−0.518
	(2.137)
Team size	0.941
	(1.400)
Perceived task interdependence in team	−0.538
	(1.223)
Prior familiarity between members	−0.582
	(0.932)
Observations	29
R^2	0.043
Adjusted R^2	−0.116
Residual Std. Error	3.854 (df = 24)
F Statistic	0.270 (df = 4; 24)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.6: Coefficients and standard errors from linear regression analysis with the average TMS score as an independent variable and team performance as a dependent variable

5.5 Research Question 3B

Does role specialization improve performance in data science teams?

I expected that, in the context of data science collaboration, having more members specialized in different functional roles will lead to higher team performance because the specialized members would be able to pool more diverse knowledge than the members from the same backgrounds. To examine this expectation, I conducted a linear regression analysis with the number of unique roles in team as an independent variable and team performance as a dependent variable. Having members specialized in certain roles did not lead to a significant increase in team performance as hypothesized as shown in Table 5.7.

	Dependent variable:
	Team performance
Intercept	22.449
	(21.777)
Number of specialized roles	3.789
	(4.128)
Team size	2.562
	(5.820)
Observations	30
R^2	0.043
Adjusted R^2	−0.028
Residual Std. Error	13.272 (df = 27)
F Statistic	0.608 (df = 2; 27)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.7: Coefficients and standard errors from linear regression analysis with the number of unique roles in team as an independent variable and team performance as a dependent variable

Chapter 6: Study 2 Methodology

6.1 Research Methods

For this study, we ran a field experiment during Data Challenge 2020 competition that took place a year after the Data Challenge 2019, which we studied in Study 1. The two events had the same formats except that they had different participants and judges. Students who participated in Data Challenge 2019 were allowed to participate in Data Challenge 2020.

6.1.1 Participants

Participants were recruited among students who voluntarily participated in the 2020 UMD Data Challenge ("Data Challenge 2020" hereafter). Participants were over 18 years of age and registered as student participants for Data Challenge 2020, hosted by the University of Maryland's (UMD) College of Information Studies. To be eligible for the study, participants were required to compete as a team of two to four members, freely selected by the participants. Of 82 teams representing the 246 students who participated in the competition, 199 students from 82 teams consented to be a part of this study. By the end of Data Challenge 2020, I collected data from

64 complete teams and 140 individuals from teams where not all team members consented to our study. If participants completed all three surveys, they were offered compensation of a $30 Amazon gift card, but they could choose whether or not to accept this compensation. This made the research inclusive to participants who may have had ethical restrictions against accepting compensation. For participants who completed only one or two surveys, we offered a $5 Amazon gift card per survey completion. Of the 199 consenting participants, 72 were female (36.2%) and 127 were male (63.8%); 87 were undergraduate students, 112 were graduate students including masters and Ph.D. students.

After Data Challenge 2020 was over, participants who were assigned to the treatment group for the field experiment were invited for in-depth interviews. To ensure that participants' memories on Data Challenge 2020 were least compromised, the interviews took place within the first 4 weeks post-Data Challenge 2020. I interviewed 18 of 58 eligible participants. A third (6, 33%) of interviewees were female and 12 were male (67%). 14 students (78%) were in graduate programs. Upon the completion of an interview, interviewees were given $25 USD in cash or as an Amazon gift card.

6.1.2 Study Design

Study 2 involved a field experiment with two conditions and employed surveys and follow-up interviews. The experiment was designed to examine the effect of developing accurate awareness of team members' expertise on building a TMS and on the effectiveness of knowledge pooling in teams. I administered four sets of surveys, one prior to the actual competition period and three during the competition. The field experiment was administered in the second survey. Teams were randomly assigned to either condition or treatment condition after the first and before the second survey. Prior to random sampling of teams, we distinguished teams into two strata, one in which all members consenting to participate in the study and the other in which partial members participating in our study. As members in teams subject to the treatment condition would be given comprehensive information about their team members, a significant information gap may have been created between teams with full and teams with partial study participation. After administering the second survey, we assessed the progression of the team process in the third and fourth surveys.

The interviews post-Data Challenge 2020 were designed to complement the field experiment. The goal of interviews was to learn how participants in the treatment group interpreted the experimental treatment (i.e., graphs showing the inaccuracies in perceived expertise in the team) and applied any findings to their collaboration. Developing such qualitative understandings could help researchers to interpret experiment results accurately. On average the interviews lasted 30 minutes.

6.1.3 Materials

Data was collected through four surveys administered before and during the Data Challenge competition: an intake survey prior to the competition period; the first-day survey on individual's knowledge/perception of other members' expertise on the first day of the competition; the experiment survey on the third or fourth day of the competition; and a final assessment survey on the last day of the competition. Judges' ratings data on the final competition submissions were shared by the Data Challenge administrators after the competition ended.

The intake survey asked about sociodemographic information such as age and gender and elicited a self-evaluation of expertise in 13 domain areas related to data science (e.g., wrangling skills, knowledge of statistics, programming skills, project management skills). The first-day survey asked about each participant's evaluation of one's own team members' expertise in these 13 domain areas.

The experiment survey was administered in two versions: one for the control condition and the other for the treatment condition. The survey for the control condition contained summarized information about Jupyter Notebook, a widely-used application for data science [98]. The survey for the experimental condition contained, along with the information on Jupyter Notebook, bar charts showing fellow team members' aggregated responses from the first survey pertaining to their perceived expertise of each member (see Figure 6.1). The visualizations illustrated how an individual member rated their expertise and how other members viewed their expertise. Both versions of surveys asked respondents to evaluate the usefulness

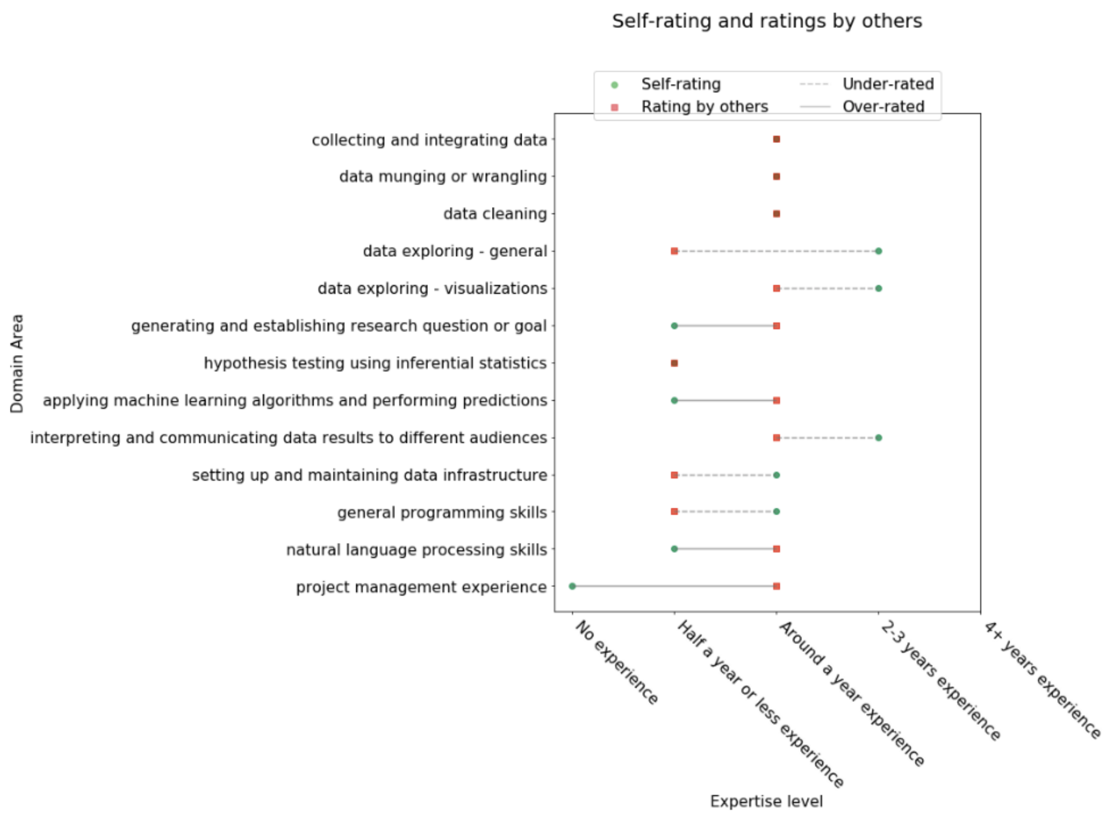

Figure 6.1: An example of a graph in the second survey representing the discrepancies in perceived expertise of team members

of the provided information and to write out what they learned from it. These questions were intended to encourage users to be more engaged with the content of a given survey.

The final survey asked student participants about the degree to which they developed collective memory on which team member knows what (i.e., a TMS); this survey mirrored Lewis's (2003) which measures behavioral indicators of a TMS. The final survey also asked how members coordinated their work during the competition ("How often did you do the following activities with your team members during the Data Challenge competition period?"); the experience of working in a team, such as how participants formed their team ("Which of the following best describes how well you knew your team members before participating in the Data Challenge?"); task division among members ("For each of the following types of activities write down which team members worked on this part of the project and what days it was worked on."); team interdependence [87]; and team satisfaction [88].

Administrative data included profiles of participants collected from Data Challenge registration forms, and judges' ratings on final projects presented on the last day of Data Challenge 2020. When student participants registered to participate in Data Challenge, they were asked to provide demographic information such as years of experience or education in a major or skills related to Data Science (e.g. statistics, visualization, design, Python libraries for data science, R, etc.). The information was used to measure the diversity of knowledge backgrounds in teams. At the end of the competition, three judges per team evaluated all final submissions made by participating teams with 33 total judges. All judges were given a rubric as well as instructions for the evaluation. The rubric measured 20 dimensions of quality, for example uniqueness and creativity of the project, the quality of final implementation or presentation, etc. For each of the dimensions judges gave a rating between 0 to 10.

6.1.4 Procedure

The first-day survey was distributed on the first day of Data Challenge 2020 (insert date here) during a kick-off event. To ensure enough time for all participants to interact with their team members, I sent out the first-day survey 6 hours after the beginning of the event (around 3pm) that day, at which point the scheduled activities of the kick-off event were completed and teams were allowed to work freely on their projects. Student participants were given a paper strip with a link to the online version (Qualtrics) of the first survey. Participants who had consented to be contacted by email and/or text were sent reminders if they missed filling out the first survey. Given the content of the experiment survey (second survey) was dependent on the responses to the first-day survey, it was imperative to allow enough time and opportunity for students to fill it out. I sent out reminders for the first-day survey a day after the kick-off event (Sunday) and early Monday morning. I stopped allowing participation in the first survey by late Monday morning.

That Monday (February 24th), two days after the first-day surveys were distributed, student participants who filled out the first-day survey received an email or text with a link to the experiment survey. After creating customized surveys for the participants in the treatment group, I sent out the experiment survey, both for the control and experiment group, around 4pm that day. The participants were contacted both the next day and the day after to remind them to fill out the experiment survey.

On the last day of Data Challenge 2020 when participants were presenting their final work and receiving evaluations, all qualifying participants received an email with a link to the final survey. Participants were given a debriefing message at the end of the final survey. I attended the final day event and reminded the participants to answer the final survey to boost response rate. Students who previously rejected to consent to our study were given opportunities to consent and fill out the final survey for partial compensation.

I then interviewed 18 participants. The first 8 interviews took place in-person. Due to the requirement to social distance and the shifting of everything online to minimize the impact of COVID-19 pandemic, the rest of the interviews were conducted through online video conferencing software. At the beginning of all interviews, interviewees were prompted to reflect on what and how they worked as a team during Data Challenge 2020. This was to revive their memories of events happened during the competition. After recalling what happened each day, I asked questions about the experimental treatment, which was the visualization of participating team members' responses. Interviewees were given a minute or two to familiarize themselves with the visualization. Once they were ready, I asked them to think aloud on what they did, in the order they did it, when they first saw the visualization, whether they thought the depiction was accurate or not, and whether any of the provided information was useful for their collaboration during Data Challenge 2020. Interviewees were thanked and debriefed on the goal of the study at the end of each interview. All interviews were recorded with permission by interviewees.

6.1.5 Data Analysis

Responses to the four surveys (i.e., intake, first-day survey, experiment survey, and final survey) were combined and cleaned for analysis. After consolidating survey responses, I dropped private information specific to these non-participants (e.g., responses to the intake survey) when only some of the members of a team agreed to participate in our research project.

6.1.5.1 Research Question 1A: Are tasks divided among members of data science teams to create differentiated roles within teams? If so, how?

For the same research question in Study 1, I did clustering analysis to see what sets of tasks co-occurred frequently and thus be clustered together based on mathematical proximity. As I identified four functional roles (i.e., generalist, research question generators, wranglers, and analysts) from the result and in order to lessen the burden on participants' having to submit daily diaries, I directly asked participants if they identified with one or more of the four roles. For example, one of the questions asked "Which of the following best describes your contribution to your team?" and respondents were given options such as a communicator and project manager, as well as the four roles.

In a real-world situation in which multiple individual members collaborate as a team, a member may assume multiple roles depending on their capabilities and the demands within a team. By letting participants select multiple roles, such cases were accounted for. In Study 1, the underlying assumption was that an individual could be characterized as only one role. One caveat is that I could not account for the roles played by the student participants who did not opt in to the study or failed to complete surveys despite consenting. However, I collected the responses from 113 student participants, about 44% of all students who participated in the event.

6.1.5.2 Research Question 1B: Is team member expertise related to which tasks someone completes?

Relative expertise in a domain area: Individual expertise was measured the same way as in Study 1 except that, in Study 2, I accounted for more dimensions of expertise related to data science. The 7 domain areas in Study 1 were expanded into 13 in Study 2 in order to capture relevant areas exhaustively. Responses were again calculated into the relative competency, same as in Study 1.

I hypothesized that individuals are likely to take up specialized roles if they have relative competency in a domain area compared to their team members. To examine this hypothesis, I fitted the data to binomial logistic regression models.

6.1.5.3 Research Question 2: Does role specialization help data science teams to develop transactive memory systems?

Independent variable

The number of unique roles in team: To capture how role specialization manifests at team-level, I aggregated the number of unique roles.

Dependent variable

The development of a TMS at the end of collaboration: I adopted a questionnaire developed by Lewis [16] to measure behavioral indicators of a TMS for individual members in a team. The individual responses were aggregated in to a team-level variable by taking an average of all individual members' responses.

Control Variable

The development of a TMS at the beginning of collaboration: I controlled for the effect of existing TMSs from members by accounting for the development of a TMS measured at the beginning of Data Challenge 2020. By controlling for this effect, the marginal effect of role specialization on the development of TMSs in teams could be accounted for.

Statistical test

I hypothesized that the more specialists that exist in teams, a lesser TMS would develop among team members. I used a multiple linear regression model to examine this hypothesis.

6.1.5.4 Research Question 3A: Does a transactive memory system improve performance in data science teams?

Independent variable

TMSs at the team level: This variable was measured in the same way as in Research Question 2.

Dependent variable

Team performance: Team performance was measured as the average score given by three judges across 20 sub-dimensions of performance (e.g., creativity, usefulness, detailed-orientedness, presentation clarity, collaboration quality, etc.). Three judges rated each team across all dimensions on the scale of 0 to 10. I took the average of the ratings across the 20 dimensions and across the three judges to generate the team-level variable.

Statistical Test

I hypothesized that more well-developed TMSs in teams would lead to higher team performance. I examined the hypothesis by fitting the data to a multiple linear regression model.

6.1.5.5 Research Question 3B: Does role specialization improve performance in data science teams?

Independent variable

Role specialization at team level: This is the same variable as role specialization at a team level in Research Question 2.

Dependent variable

Team performance: This is the same variable as team performance in Research Question 3A.

Statistical Test

I hypothesized that the more specialists that exist in a team, the better the team performance would become. I fitted the data to a linear regression model to test this hypothesis.

6.1.5.6 Research Question 4A: Does highlighting inaccuracies in perceptions of team member expertise improve the development of transactive memory systems in data science teams?

6.1.5.7 Research Question 4B: Does highlighting inaccuracies in perceptions of team member expertise increase the amount of explicit coordination in data science teams?

6.1.5.8 Research Question 4C: Does highlighting inaccuracies in perceptions of team member expertise improve team performance in data science teams?

I hypothesized that highlighting inaccuracies in perceived expertise of team members would improve the development of a TMS, increase explicit coordination, and improve team performance. I tested each question using a multiple linear regression with the treatment assignment as a binary variable.

Interview recordings were transcribed using automatic transcribing software, Temi (`temi.com`). The transcriptions were compared with the jottings I took during each interview and corrected for any typos and inaccurate transcriptions. Transcribed interview data were open-coded to answer research question 4A, B, and C.

Chapter 7: Study 2 Results

Four research objectives were addressed in Study 2 in the form of four research questions divided into eight sub-questions. The main goal of the study was to test if highlighting inaccuracies in perceived expertise would benefit data science teams by enhancing the development of a TMS as well as explicit coordination. I conducted a field experiment to test the hypothesis and the results are discussed across Research question 4A, B, and C. Research questions 1 through 3 inherited almost the same constructs and hypothesis testing methods as Study 1 to make sure that the same observations would be replicated in Study 2 context.

7.1 Research Question 1A

Are roles divided among members of data science teams? If so, how?

As in Study 1, I expected that data science teams would be composed of members taking on varying task roles. Participants in Study 2 who answered the question asking what roles they played came from 80 different teams (out of a total of 82) participating in Data Challenge 2020. Individuals could select all of the roles they played during the competition period.

The role played most frequently selected by the respondents was analyst

Role name	Count (%)
Analyst	71 (62.8)
Wrangler	53 (46.9)
Presenter	51 (45.1)
Generalist	49 (43.4)
Research Question generator	49 (43.4)
Communicator	48 (42.5)
Project manager	48 (42.5)
Domain expert	33 (29.2)

Table 7.1: The number of individual members taking on each role

	Project Manager	Communicator	Wrangler	Analyst	Domain Expert	RQ* Generator
Project Manager	.	.	.	.	.	.
Communicator	31	.	.	.	.	.
Wrangler	28	21	.	.	.	.
Analyst	34	32	38	.	.	.
Domain Expert	20	16	18	27	.	.
RQ Generator	29	27	26	30	19	.
Presenter	27	28	25	35	21	31

*: Abbreviation of "research question"

Figure 7.1: Co-occurrence matrix of roles at an individual level

(71/113, 63%) followed by wrangler (53, 47%) and presenter (51, 43%). The least played role was the domain knowledge expert (33, 29%). Previous studies focused on data science teams in corporate settings [23, 47, 48] and studied mostly technical roles such as data engineers, data scientists, or machine learning engineers. The results from Study 2 show that soft roles are as prevalent as technical roles in data science teams. The distribution across all roles is shown in Table 7.1.

Figure 7.1 shows how often individuals took on a combination of two roles. At individual level, the most commonly assumed pairs of roles are analyst-wrangler (38 co-occurrences), analyst-presenter (35), and analyst-project manager (34).

In 29 out of 80 teams, all members answered the role question. Of those 29 teams, the median team size was 4 and mean team size was 3.24. The median

	Project Manager	Communicator	Wrangler	Analyst	Domain Expert	RQ Generator
Project Manager	.	.	.	.	.	.
Communicator	30	.	.	.	.	.
Wrangler	31	32	.	.	.	.
Analyst	34	35	37	.	.	.
Domain Expert	18	19	21	24	.	.
RQ Generator	29	30	27	30	18	.
Presenter	29	33	30	32	18	29

Figure 7.2: Co-occurrence matrix of roles at a team level

number of the roles in teams was 6 with mean of 5.5. The mean number of roles per team size varied slightly (2 person team: 5.6, 3: 5.7, 4: 5.6). This study is one of only a few that shows individual members in data science teams play multiple roles, often overlapping with each other. Figure 7.2 shows how often teams had a combination of two roles assumed by any members in each team. Teams most commonly had a combination of analyst-project manager (34), analyst-communicator (35), and analyst-wrangler (37).

7.2 Research Question 1B

Is team member expertise related to which tasks someone completes?

I expected that members of data science teams would take on task roles that were closely related to their past experiences and expertise. To examine this hypothesis, I conducted logistic regression analyses between the 13 domains of expertise related to data science work as independent variables and an assumption of a role as a dependent variable. All expertise variables were included in each of the logistic regression models as none of the variables had a collinearity issue (i.e., VIF score ranging from 1.4 to 3.3).

To understand the association between relative competency and role assumption behavior, I centered experience data to the mean of each team and accounted for team-level differences in the statistical model. Table 7.2 and Table 7.3 show the results of fitting a mixed-effect logistic model for each role variable as a dependent variable and 13 variables measuring expertise in different domains as predictors.

For each additional year of experience in research question development compared to the average of an individual's team members increased the odds of assuming a project manager role by 94% at a marginally significant level (coef. = 0.68, $p = 0.08$). Relative competency in data wrangling and programming compared to team members lowered the odds of assuming a communicator role by 76% ($p < 0.01$) and 53% respectively ($p = 0.02$). On the contrary, each additional year of experience in visualization and project management increased the odds of assuming a communicator role by 154% ($p = 0.04$) and by 75% with a marginal significance ($p = 0.07$) respectively.

The relative competency in programming compared to the average of other team members was linked to higher odds of assuming a wrangler role by 115% ($p = 0.02$) and an analyst role by 112% ($p = 0.03$), but lowered the odds for assuming a generalist role by 47% ($p = 0.05$). Relative competency in interpretation skills lowered the odds of assuming an analyst role by 49% ($p = 0.03$). For a generator role, relative competency in data infrastructure management skills increased the odds of role assumption (103%, $p = 0.05$), yet competency in data collection (48%, $p = 0.06$), visualization (50%, $p = 0.09$), and programming (47%, $p = 0.05$) were marginally linked to lowered odds. Relative competency in data exploration was associated with lower odds of taking on a presenter role (62%, $p = 0.04$). No

	Dependent variable:			
	Project manager	Communicator	Wrangler	Analyst
Intercept	-0.359^{*}	-0.413^{*}	-0.143	0.674^{***}
	(0.206)	(0.226)	(0.207)	(0.246)
YOE: Data collection	-0.069	-0.090	-0.065	0.247
	(0.324)	(0.360)	(0.335)	(0.344)
YOE: Data munging	-0.703^{*}	-1.427^{***}	0.420	0.122
	(0.379)	(0.468)	(0.376)	(0.409)
YOE: Data cleaning	0.638	0.288	-0.038	0.062
	(0.434)	(0.496)	(0.425)	(0.467)
YOE: Data exploration	-0.170	-0.148	0.291	-0.015
	(0.392)	(0.446)	(0.394)	(0.417)
YOE: Visualization	-0.017	0.936^{**}	-0.467	0.704
	(0.393)	(0.456)	(0.396)	(0.468)
YOE: RQ development	0.678^{*}	0.199	-0.003	0.488
	(0.393)	(0.429)	(0.397)	(0.427)
YOE: Inferential statistics	-0.033	0.085	0.200	0.077
	(0.360)	(0.422)	(0.375)	(0.407)
YOE: Predictive modeling	0.107	-0.526	-0.256	-0.087
	(0.367)	(0.420)	(0.400)	(0.422)
YOE: Interpretation of analysis	-0.093	0.311	-0.179	-0.668^{**}
	(0.249)	(0.282)	(0.258)	(0.315)
YOE: Data infra management	0.224	0.554	-0.078	-0.629
	(0.334)	(0.382)	(0.357)	(0.383)
YOE: Programming	-0.397	-0.756^{**}	0.765^{**}	0.750^{**}
	(0.300)	(0.342)	(0.323)	(0.339)
YOE: Natural Language Processing	-0.197	0.368	-0.179	0.738
	(0.409)	(0.456)	(0.408)	(0.490)
YOE: Project management	0.146	0.563^{*}	-0.462	0.004
	(0.269)	(0.307)	(0.294)	(0.293)
Observations	113	113	113	113
Log Likelihood	-71.118	-61.114	-70.355	-63.401
Akaike Inf. Crit.	172.235	152.228	170.709	156.801
Bayesian Inf. Crit.	213.146	193.139	211.620	197.712

YOE: An abbreviation of years of experience

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7.2: Coefficients and standard errors from logistic regression models with relative competency in data-science related skills compared to team members as independent variables and self-reported task roles as a dependent variable - part 1

	Dependent variable:			
	Domain expert	RQ generator	Generalist	Presenter
Intercept	-2.197^{**}	-0.305	-0.388^{*}	-0.308
	(1.114)	(0.215)	(0.212)	(0.262)
YOE: Data collection	-0.222	-0.553^{*}	-0.651^{*}	0.291
	(0.613)	(0.335)	(0.348)	(0.357)
YOE: Data munging	-0.220	-0.308	-0.054	-0.025
	(0.712)	(0.374)	(0.378)	(0.408)
YOE: Data cleaning	-1.166	-0.026	0.436	0.223
	(0.865)	(0.420)	(0.425)	(0.473)
YOE: Data exploration	0.517	0.431	0.165	-0.956^{**}
	(0.768)	(0.398)	(0.391)	(0.465)
YOE: Visualization	-0.762	-0.749^{*}	-0.701^{*}	0.148
	(0.834)	(0.416)	(0.418)	(0.438)
YOE: RQ development	0.963	0.196	0.323	0.659
	(0.723)	(0.385)	(0.390)	(0.447)
YOE: Inferential statistics	0.252	0.142	-0.587	-0.564
	(0.700)	(0.360)	(0.383)	(0.426)
YOE: Predictive modeling	1.112	0.183	0.147	0.348
	(0.906)	(0.377)	(0.388)	(0.414)
YOE: Interpretation of analysis	0.685	0.229	-0.043	-0.074
	(0.556)	(0.257)	(0.258)	(0.277)
YOE: Data infra management	0.574	0.335	0.710^{**}	0.059
	(0.692)	(0.339)	(0.357)	(0.379)
YOE: Programming	0.193	-0.144	-0.641^{**}	-0.048
	(0.558)	(0.298)	(0.321)	(0.330)
YOE: Natural Language Processing	0.357	0.257	0.189	-0.188
	(0.788)	(0.424)	(0.420)	(0.455)
YOE: Project management	-0.174	0.160	0.306	0.376
	(0.520)	(0.272)	(0.278)	(0.308)
Observations	113	113	113	113
Log Likelihood	-54.770	-72.387	-69.388	-71.594
Akaike Inf. Crit.	139.540	174.775	168.775	173.188
Bayesian Inf. Crit.	180.450	215.686	209.686	214.099
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 7.3: Coefficients and standard errors from logistic regression models with relative competency in data-science related skills compared to team members as independent variables and self-reported task roles as a dependent variable - part 2

predictors were significantly associated with taking on either of domain expert or research question generator role.

7.3 Research Question 2

Does role specialization help data science teams to develop transactive memory systems?

This research question is identical to that in Study 1, but the I defined the statistical model to include average TMS scores at the beginning of the collaboration to account for any confounds. As in Study 1, I constructed a multiple linear regression model with the number of unique roles in teams as an independent variable and the average TMS score measured in the last stage of collaboration as a dependent variable. Table 7.4 shows the result of this linear regression analysis with the number of unique roles in teams as an independent variable and the average value of all team members' responses to 15 variables measuring the three dimensions of TMS manifestation as a dependent variable. I compared the two models with and without the average TMS score measured at the beginning of Data Challenge 2020 as a confound.

The number of unique roles per team was marginally linked to increased TMS development by the end of the collaboration period (coef. $= 0.06$, $p = 0.07$) as shown in Table 7.4. However, when accounting for the level of TMS development measured at the beginning of the collaboration, the association of role specialization and TMS development by a later stage was insignificant (coef. $= 0.03$, $p = 0.3$).

TMS development at the beginning was statistically associated with TMS development at the later stage (coef. $= 0.54$, $p < 0.01$), indicating that the association between role specialization at a team level and TMS development was confounded by the development of TMS at the beginning of collaboration. TMS at the beginning of collaboration explained 24% of variance in TMS development by the final stage of collaboration. The unique number of roles in team alone explained only about 6.4% of the variance of the TMS measured at the final stage of collaboration.

A correlation analysis between the number of unique roles per team and early TMS showed that the two constructs were significantly correlated though the explained variance was small (coef. = 1.59, p = 0.02, $R^2 = 0.09$). As teams delegated and worked on tasks after the early TMS was formed, this result indicated that the teams with relatively high TMSs from the beginning tended to have more unique task roles with which the members self-identified in the end compared to the teams with relatively low TMSs, and their early TMSs stayed intact until the end of the collaboration as implied by their high final TMS scores.

The post hoc correlation analysis between early TMS development and final TMS development showed that the two variables were highly correlated (coef. $=$ 0.54, $p < 0.01$)

7.4 Research Question 3A

Does a transactive memory system improve performance in data science teams?

Literature on the association between TMSs in teams and team performance

	Dependent variable:	
	TMS in the end	
Intercept	3.501^{***}	1.580^{***}
	(0.175)	(0.491)
TMS in the beginning		0.541^{***}
		(0.131)
Total number of unique roles in team	0.056^{*}	0.028
	(0.030)	(0.027)
Observations	52	52
R^2	0.064	0.304
Adjusted R^2	0.045	0.276
Residual Std. Error	0.430 (df = 50)	0.375 (df = 49)
F Statistic	3.409^{*} (df = 1; 50)	10.725^{***} (df = 2; 49)
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 7.4: Coefficients and standard errors of a multiple linear regression analysis with the number of unique roles in teams as an independent variable and TMS score measured at the final stage of collaboration as a dependent variable

suggest the association may vary depending on the type of tasks and how performance is measured [50, 97, 99]. To examine the association in the context of data science collaboration, I conducted a multiple linear regression analysis with the average TMS score measured at the beginning and at the last stage of collaboration as main predictors and team performance as a dependent variable. One caveat of the analysis was that the consistency between the raters was poor (Intra-class correlation coefficient value of 0.23) [100].

The extent of TMS development measured at either the beginning or the last stage of collaboration were not associated with team performance (See Table 7.5). The results showed no association between either early or final TMS scores and team performance.

I further analyzed the association between each sub-dimension of 20 variables that constituted the average team performance and final TMS in team. Final TMS score in team explained more variance in some of the sub-dimensions compared to the average team performance, but the magnitude of effect was marginal at best. For example, the model with a performance sub-dimension measure on how much additional information has been provided by the final project outcome as a dependent variable had R^2 value of 0.07. The model showed a marginal association between both measures of TMS and the sub-dimension team performance [1].

7.5 Research Question 3B

Does role specialization improve performance in data science teams?

[1] Final TMS: Coef. = 1.43, p = 0.08, Early TMS: Coef. = -1.04, p = 0.16

	Dependent variable:	
	Team performance	
Intercept	7.970***	6.982***
	(1.492)	(1.705)
Early TMS	0.089	
	(0.301)	
Final TMS		0.327
		(0.374)
Team size	−0.134	−0.115
	(0.206)	(0.204)
Observations	46	45
R^2	0.014	0.028
Adjusted R^2	−0.032	−0.018
Residual Std. Error	0.933 (df = 43)	0.934 (df = 42)
F Statistic	0.308 (df = 2; 43)	0.606 (df = 2; 42)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 7.5: Coefficients and standard errors of two multiple linear regression models with the average TMS scores in teams measured in the beginning and in the end as an independent variable respectively, and team performance as a dependent variable

The existence of members taking on certain team roles can ensure efficient communication and management of available resources in team collaboration [36,43]. As data science teams pool from various fields of knowledge, I predicted a strong tie between the coverage of unique roles in teams and team performance and tested this by conducting a multiple linear regression analysis with the number of unique roles in teams as an independent variable and team performance as a dependent variable. The number of unique roles was marginally associated with the team performance (coef. $= 1.62$, $p = 0.09$) as shown in Table 7.6. This means that one unit increase in the average specialization of roles in team members significantly led to a one point increase in team performance. The model explained the 6% of the variance in team performance ratings.

	Dependent variable:
	Team performance
Intercept	6.419***
	(0.841)
Number of specialists	1.617*
	(0.930)
Observations	47
R^2	0.063
Adjusted R^2	0.042
Residual Std. Error	0.964 (df = 45)
F Statistic	3.024* (df = 1; 45)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7.6: Coefficients and standard errors of a multiple linear regression analysis with the number of unique roles in teams as an independent variable and team performance as a dependent variable

7.6 Research Question 4A, 4B, and 4C

By addressing research questions 4A, B, and C, I assessed if highlighting inaccuracies in perceived expertise in teams would be beneficial to data science teams.

Based on discrepancies between responses on self-evaluations and evaluations made by other team members, the top three domain areas in which participants rated themselves most competently were *generating and establishing research question or goal* (mean value of 3.16, meaning between 2-3 years of experience), *general programming skills* (3.11), and *data exploration skills through visualization* (3.07). The top three areas in which the average ratings by other team members were the highest were *interpreting and communicating data results to different audiences* (3.33), *data exploration skills through visualization* (3.18), and *general programming skills* (3.16).

The domain areas in which the evaluations by the other team members underrated than self-evaluations the most were natural language processing skills (0.63, meaning of underrating than self-evaluations by less than a year), skills related to collecting and integrating data (0.62), and applying machine learning algorithms and performing predictions (0.41).

The most overrated domain areas by other team members were interpreting and communicating data results to different audiences (-0.31), data exploration skills through visualizations (-0.14), and general data exploration skills (-0.13).

4A: Does highlighting inaccuracies in perceptions of team member expertise improve the development of transactive memory systems in data science teams?

The mean values of final TMS scores were 3.76 for the control and 3.86 for the treatment group (standard deviations were 0.42 and 0.46 respectively). Yet, there was no significant association between highlighting inaccuracies in perceived expertise and final TMS scores as shown in Table 7.7[2].

	Dependent variable:
	TMS in team
Intercept	1.587***
	(0.494)
TMS in the beginning	0.570***
	(0.128)
Highlighting inaccuracies	0.078
	(0.105)
Observations	52
R^2	0.297
Adjusted R^2	0.268
Residual Std. Error	0.377 (df = 49)
F Statistic	10.359*** (df = 2; 49)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7.7: Coefficients and standard errors of a linear regression model with highlighting inaccuracies in perceived expertise as an independent variable and final TMS development as a dependent variable

4B: Does highlighting inaccuracies in perceptions of team member expertise increase the amount of explicit coordination in data science teams?

The mean values of how often members discussed their expertise in teams were 3.49 for the control[3] and 3.34 for the treatment group (standard deviations were 0.81

[2]Further analyses with each of the three sub-dimensions of TMSs showed insignificant association with the experimental intervention.

[3]The value of 3 indicated that they discussed once in a couple days during the competition. 4

and 0.88 respectively). The mean values of the frequency of explicit coordination were 3.22 for the control and 3 for the treatment group (standard deviations were 0.62 and 0.53 respectively). The mean values of the perceived quality of explicit coordination were 4.21 for the control and 4.23 for the treatment group (standard deviations were 0.81 and 0.65 respectively). Overall, groups in the experiment condition discussed expertise less and engaged in fewer coordination activities, but they rated higher in their quality of coordination compared to the control groups. Yet there was no significant association between highlighting inaccuracies in perceived expertise and the frequency of expertise discussion, explicit coordination, or the quality of explicit coordination as shown in Table 7.8.

	Dependent variable:		
	Expertise discussion	Frequency of coordination	Quality of coordination
Intercept	3.491***	3.223***	4.209***
	(0.162)	(0.111)	(0.143)
Highlighting inaccuracies	−0.147	−0.220	0.023
	(0.234)	(0.161)	(0.206)
Observations	52	52	52
R^2	0.008	0.036	0.0003
Adjusted R^2	−0.012	0.017	−0.020
Residual Std. Error (df = 50)	0.842	0.579	0.742
F Statistic (df = 1; 50)	0.398	1.880	0.013
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7.8: Coefficients and standard errors of three linear regression models with highlighting inaccuracies in perceived expertise as an independent variable and each of three measures of explicit coordination as a dependent variable

4C: Does highlighting inaccuracies in perceptions of team member expertise improve team performance in data science teams?

indicated that they discussed once or twice a day.

The mean values of final TMS scores were 4.21 for the control and 4.23 for the treatment group (standard deviations were 0.81 and 0.65 respectively). There was no significant association between highlighting inaccuracies in perceived expertise and final TMS scores as shown in Table 7.7.

	Dependent variable:
	Team performance
Intercept	7.760***
	(0.173)
Highlighting inaccuracies	0.144
	(0.266)
height Observations	54
R^2	0.006
Adjusted R^2	−0.013
Residual Std. Error	0.966 (df = 52)
F Statistic	0.294 (df = 1; 52)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7.9: Coefficients and standard errors of a linear regression model with highlighting inaccuracies in perceived expertise as an independent variable and team performance as a dependent variable

Highlighting inaccuracies in perceived expertise of team members did not significantly affect final TMS, quantity and quality of explicit coordination, and performance in teams. Two explanations are possible for the results. One, the intervention didn't incur the intended effect in teams. Two, the intervention may have incurred the effect but it was not detected due to the low statistical power used to examine the hypothesis.

The data on the duration data on how long participants took to read the experiment survey supported the first explanation. It is expected that participants in the treatment group would spend more time reading the experiment survey because the survey contained the graphs that were not included in the control group's survey. Yet, no difference in the minutes spent reading the report was found between the treatment and control groups. After removing two outliers, 277 and 122 minutes in each group respectively, treatment group spent average 7.7 minutes while control group spent 7.4 minutes. T-test result also indicated that there was no difference in average minutes the groups spent to read the report between the two groups ($p = 0.8$). In short, participants in the treatment group may have not carefully read the experiment survey enough for them to update their perception with provided information. This could have resulted in no difference in multiple measures on team perception (i.e., TMS), coordination, and performance.

To the same end, no significant difference in the perceived usefulness of the content of the survey was found between the control and treatment group with a mean rating of 3.13 [4] for the control and 3.17 for the experiment group ($p = 0.8$)

Although there were no significant mean differences in how useful individuals found the intervention, a few individuals reported finding the intervention helpful[5]. The participants mentioned that they would use the provided information to allocate team members to tasks matching their expertise, to adjust self-attitude when collaborating, and to trust their team members in the areas of their expertise.

Know that I have a better understanding of the level of experience my

[4]The rating value of 3 corresponded with *moderately useful*, and 4 corresponded with *very useful*.

[5]21 of 58 responses from the treatment group who rated "very useful" or "extremely useful"

> group perceives both themselves and me, I can better steer the group to a developing certain skills. For instance, if they listed something as being lower, I can task them with developing some more of those skills.

> This information will boost my confidence and allow me to converse with my team members in a more confident manner.

> I can rely on my team members regarding some aspects they're comfortable with.

10 of 58 responses from the treatment group rated the usefulness of the content to be "slightly useful" or "not at all useful". They mentioned that they had doubts for self-ratings, and the content did not add anything new.

7.6.1 Post-experiment interviews

After Data Challenge 2020 was over, I conducted post-experiment interviews with 18 student participants whose teams were in the treatment group. The goal was to gain a deeper understanding of how they, as individuals and as a team, interpreted highlighted inaccuracies in perceived expertise within teams.

Some interviewees reported that they could reflect on the domain areas they should develop. For example:

I noticed that I have little expertise compared to others; I came to think more about what I can do for the team, what I should improve after or during this data challenge.(P2)

The part about me was useful because that's related to how teams would evaluate me. It's better to know where you stand, right?(P4)

(Identifying inaccuracies in perceived expertise) was good for self-examination to know where to improve and find blind spots.(P10)

Highlighting inaccuracies of perceived expertise in sub-fields of data science helped some teams discover previously unknown expertise available in the teams:

The good thing about this graph is it breaks down into more skill sets. (...) We were able to ask our members to do more on this area based on these graphs.(P10)

(If this content had been shown earlier) it might actually change my perspectives. I might look at it and "hey, this person is not really strong in X so we might do something different".(P1)

The interviews also revealed several issues that could explain the field experiment results. First, the majority of participants had relatively limited expertise so there were no large enough discrepancies to affect collaboration dynamics within teams. Participants had the least experience in natural language processing (NLP) with less than one year of experience on average. They had the most experience in programming with an average of just over two years of experience. Consequently, the magnitude of discrepancy was mostly capped within this limited range of experiences—two years at most. Given that one or two years of experience in a domain area is considered beginner level experience, knowing whether a person had 2 years of experience when one had thought for the person to have only a year of experience was too trivial to alter collaboration dynamics.

"I am not surprised to see the discrepancies between my ratings versus team

member's ratings because what counts as experience is different. For example, if you worked for a company for a year as a data scientist would you consider that a year experience in data cleaning? (..) So I don't trust a year or two of difference in ratings." (P3)

Second, interviewees from some teams did not have much faith in the discrepancy information provided as a report (i.e., the experimental intervention) due to the timing of receipt. The report was generated based on the survey responses collected on the first day of Data Challenge 2020 and members in the teams did not know much about their team members because they did minimal work on the first day. They responded to the first-day survey based on almost random guess on their team members' expertise, leading them to distrust the content of the report.

"We've known each other for a long time so we laughed off of the discrepancies in ratings." (P3)

"I rated slightly useful because at that time we didn't know each other very well; so it's not based on fact, it's based on impressions." (P5)

"I think one week is not enough to get to know the person so maybe the result was not that accurate." (P6)

Lastly, the duration of the competition was too short for teams to reflect on what they discovered from reading the discrepancy information. Most of the interviewees reported that they worked for three to four days during Data Challenge 2020. As all teams were required to participate in a kick-off event on the first day, this means that teams had around two to three days to collaborate after reading the experimental intervention.

"It (The competition period) is too short. If you had a month to do the project it would help, but since it was for a week what would you do with three days?" (P4)

"We didn't start working until Thursday. Most of the work was done in Friday. If it was sent out during that day it might give a different picture." (P10)

Chapter 8: Discussion

The goal of this study was to facilitate group memory pertaining to available expertise and its location (i.e., a transactive memory system, or TMS) in data science teams. To achieve this goal, I set out four objectives and eight different research questions across two studies. The objectives are: (1) To characterize role specialization in data science teams and whether it is based on expertise, (2) to understand the relationship between role specialization and transactive memory systems in real-world data science teams, (3) to understand whether either role specialization or forming a transactive memory system helps data science teams perform better, and (4) to understand if highlighting inaccuracies facilitates explicit coordination, the development of transactive memory systems, and improves team performance in data science teams. The first study aimed to achieve the first three objectives by collecting and analyzing observational data on a university-wide data science competition, Data Challenge 2019. The second study attempted to reaffirm the first three objectives as well as to achieve the fourth objective. In the second study, I conducted a field experiment and collected data through surveys and interviews in Data Challenge 2020. I discuss the findings for each research objective in the following sections.

8.1 Research Objective 1: To characterize role specialization in data science teams and whether it is based on expertise

In this research, roles taken by individuals in data science teams were defined at a more granular level than previous studies on data science collaboration [23,47,48]. I investigated the taxonomy of roles in two ways: based on clustering analysis results in Study 1 and based on self-reports in Study 2. The role measurement in Study 2 (self-select) changed from that in Study 1 (clustering analysis) in order to account for how members took on non-technical roles in addition to the essential roles identified in Study 1. Additionally, through self-selecting roles, participants could choose more than one role they did in team work. Study participants reported that they worked as generalists less frequently compared to the ratio of generalists identified in the clustering analysis of their task records. According to the results of the clustering analysis, the majority of the participants (82 of 118, 69.5%) were categorized as generalists in Data Challenge 2019. In Data Challenge 2020, fewer participants (49 of 113, 43%) self-reported that they worked as generalists.

The gap in the percentage of generalists between the two studies can be attributed to two reasons. One, the notion of generalist was more broadly construed in Study 1 than in Study 2. Generalists in Study 1 were those who were not specialists; specialists were someone who devoted most of their time to only one category of tasks. In Study 2, participants were allowed to identify themselves with any role(s) out of eight options, an expanded list from that in Study 1. The definition of the generalist role in Study 2 was operationalized to be “someone contributing to all tasks partially rather than having clearly defined roles.”

Two, the substantial shift in the proportion of generalists between both studies indicates that role identification is subject to the measurement method employed. To the best of my knowledge, consensus on the suitable measurement method for role identification has yet to be reached across literature on team roles. In studies on team role theory, Belbin and colleagues used two methods to identify individual members' naturally-emerging roles in teams: through self-reports and observations by team members [42, 43]. More scalable methods of identifying roles than these methods include doing clustering analysis based on activity traces [93, 101], which was adopted in Study 1.

The distribution across specialized roles was roughly even in each study. In Study 1, the percentages of research question generators, wranglers, and analysts were 11.9% (14 of 118), 9.3% (11), and 9.3% (11), respectively. When individuals self-identified their role in Study 2, the roles were roughly evenly split with two exceptions: 48% of the participants self-identified as the project manager and the communicator, respectively; 49% self-identified with the research question generator and the generalist role, respectively; 51% self-identified as the presenter; and 53% self-identified as the wrangler. A much high percentage self-identified as an analyst (71 of 113, 62.8%). A much lower percentage self-identified a domain expert (33 of 113, 29.2%).

Different role identification methods used led to different role representations in Study 1 and Study 2. I would argue that the representation shown in Study 2 is a more accurate representation of the real-world data science collaboration, that is, an individual member takes on multiple task roles in any collaboration situations [102]. One interesting finding of Study 2 is that some roles often overlooked in research on data science collaboration, such as project managers[1] and communicators[2], had a similar representation as the other roles requiring technical knowledge, such as wrangler and analyst. One suggestion based on the findings is that when studying data science teams, researchers should account for non-technical roles (i.e., project manager, communicator, or presenter) as well as technical roles (i.e., analyst or wrangler).

Although individuals identified with different types of roles, the combination of roles observed in teams barely varied. Analyst and wrangler could be considered as the essential roles for data science teams as those two co-appeared most frequently in both studies, though the type of roles in Study was limited to only four roles. The second and third most co-appearing roles with the analysts were communicators and project managers. According to the individual level co-occurrence matrix, individual members often take on the combinations of roles at the same time. This suggests that members who were engaged in the analysis tasks likely acted as a hub of their teams by being engaged in communications between members and management of the work.

While I did not find any associations between roles and expertise in Study 1, I was able to find associations in Study 2. I found associations between most roles

[1] Project manager: Someone taking leadership and organizing project timeline and details
[2] Communicator: Someone facilitating communication between team members

and expertise in skills required for those roles. For example, I found an association between having programming skills and increased odds of taking on the analyst and the wrangler roles respectively. Being relatively competent in programming skills was also associated with decreased odds of taking on the communicator and the generalist roles.

It is worth noting that in Study 2, non-technical roles such as project manager, communicator, or presenter roles were negatively associated with having experiences in technical domains. For example, having experience in data wrangling, predictive modeling, and programming were linked to lower odds of taking on the communicator role. On the contrary, technical roles, such as wrangler and analyst, were not associated with having experience in non-technical domains such as project management or research question development. This pattern was consistent across two sets of analyses, one set with expertise measured in years and the other set with relative competency of one's expertise in comparison to team members. This suggests that assuming technical roles is about the presence of skills, whereas taking on non-technical roles is about the absence of technical skills.

I argue that the association between expertise in roles found in Study 2 is more accurate than the lack of association found in Study 1. I believe that measurement of roles was better in Study 2 because Study 2 included more non-technical roles than Study 1 such as communicator, presenter, and research question generator and doing so could break down what was previously lumped into as the generalist role in Study 1. Also, the self-report method used in Study 2 to identify roles was better at capturing natural role assumptions than Study 1 by allowing the assumption of multiple roles. This is left for future research, because I cannot rule out the possibility that the differences were due to differences between Data Challenges 2019 and 2020. The mismatch between Study 1 and 2 may come from the fact that the clustering analysis in Study 1 failed to accurately represent who took on which roles in data science teams. Roles other than the generalist could have been identified too rarely, and subsequently the power to detect any significant association between expertise and association may have been too small with only few of the roles identified. Or it could be that more experienced students partook in Data Challenge 2020 than in Data Challenge 2019 so the members in this year's event were more prone to elect roles that matched with their expertise.

8.2 Research Objective 2: To understand the relationship between role specialization and transactive memory systems in data science teams

In summary, I found no association between the degree of role specialization in a team and the development of a TMS at the last stage of collaboration. There are at least two possible explanations for why I did not find an association between the two. One, I may not have had enough power to detect an association due to the small sample number of groups in each study. For example, in Study 2 there was 35% chance to detect the association between the independent and dependent variables at 0.05 significance level.

Two, a TMS may have formed early in the teams, when it did form, independent of what happened during the collaboration. In Study 2, All teams agreed that their teams exhibited the behavioral indicators of a TMS from the beginning of the collaboration. The ceiling effect from the high ratings for early TMS could have made it even harder to detect any changes in the development of a TMS during the Data Challenge.

If there was indeed no association between the number of specialized roles and TMS development in the final stage of collaboration, it may suggest that how teams divide their work may not affect the development of TMS in cross-functional teams. One possible explanation for this is that individual members form their own expectations about other members' abilities and the individuals hardly change the expectations because they are unfamiliar with the others' expertise areas.

8.3 Research Objective 3: To understand whether role specialization or forming a transactive memory system helps data science teams perform better

In two analyses, performed in both Study 1 and 2, the development of a TMS in a team was not found to be associated with team performance. The power of the model to detect any association between the two variables was 15.4% (degree of freedom = 3, sample size = 42, $R^2 = 0.04$). This finding is not consistent with previous studies implying an association between a well-developed TMS and high team performance [96,97,103–105], though team performance has been known to be contextualized to a given task goal so the association between the two has not been always identified. This mismatch may indicate that an understudied mechanism may exist that transfers the benefits of a well-developed TMS to enhance team performance, but the mechanism was not at work for the data science teams in Data Challenge 2019 and 2020.

Studies on the development of TMSs in cross-functional or multidisciplinary teams in which members have different functional backgrounds have identified varying effects of cross-functionality to the development of a TMS. Kotlarsky and colleagues claimed that a highly-developed TMS in a team at the beginning of the collaboration helped to reduce the knowledge boundaries in cross-functional teams [106], but knowledge boundaries could hinder the further development of TMSs in teams during team work [107]. The model explained 41% variance of the TMS development. Liao and colleagues found the mediating role of team identification bridged the effect of communication quality on TMS development [108]. In other words, the more you, as a team member, felt the communication quality in your team satisfactory, the more you would identify with the team, thus facilitating the development of a TMS. The explanatory power of the model used for hypothesis testing was not disclosed. However, those studies did not account for the relationship between the development of TMSs and team performance in such team settings [106–108]. To the best of my knowledge, this study is the first attempt to research the development of TMSs specifically in cross-functional teams and its impact on team performance.

Overall there was inconclusive support for the association between the number of unique roles in team and team performance. Study 1 model showed no association between the two constructs. Study 2 model showed marginal support for the association, but it was found to be due to the confound of early TMS. Both models had low statistical powers to detect any underlying association (15.2% and 17.5% respectively).

When no association between the number of unique roles and team performance is assumed, it could be explained in two ways. Firstly, because of high early TMS scores in most of the teams, the number of roles did not have any effect on team processes that may be critical to the team performance, but not captured by this study. This could have led to no effect on team performance. Secondly, the roles defined in this study could have been irrelevant to team performance in data science collaboration. Some roles defined in this study matched the nine team roles identified by Belbin in his Team Role Theory [42]: Project manager as shaper, communicator as coordinator, research question generator as plant, and domain expert as resource investigator. Analyst, wrangler, and presenter all may be considered specialists. The other four team roles, monitor-evaluator, team worker, implementer, and completer-finisher, were not captured in our study because adopting the all nine roles as well as functional roles tailored to a data science context would have resulted in a too long a list of roles and could have dropped the survey response rate. Even with the 5 roles accounted for in the study, the results showed that teams with more coverage of those roles did not lead to higher team performance. This may indicate that team performance in data science teams may not be tied to the coverage of team roles.

8.4 Research Objective 4: To understand if highlighting inaccuracies in perceptions of team member expertise facilitates the development of transactive memory system, the amount of explicit coordination, and team performance in data science teams

On average, teams that were given a report containing information highlighting inaccuracies in their perceptions of team member expertise discussed expertise less often and engaged in fewer coordination activities, but rated their quality of coordination higher when compared to control groups. These teams had higher team performance scores than teams in the control group. Yet, the differences in means were not statistically significant to be regarded as real effects of the treatment. Assuming that highlighting inaccuracies in perceived expertise failed to incur the intended effects within teams, several factors that could explain this result were discovered in the post-competition interviews.

One, the life cycle of data science projects could have affected the interpretation of reports. I administered the survey on the third day of Data Challenge 2020 and, by that time, most participants were working on data cleaning. So, if you had good skills in cleaning and wrangling you would have been overrated overall. If your skills would be used in a later stage (e.g., predictive modelling and inferential stats), then other team members would not be as aware of those skills. Therefore, the duration of collaboration could have been too short to apply anything learned from the report. Students did not work on the data challenge whole time during the competition; some of them had midterms, or other academic schedules. As a result, most of the interviewees reported that they had worked on the data challenge for about three to four days. It was impractical for the participants to reflect on the newly-earned knowledge from the experimental survey to alter their team coordination and workflow. Lastly, the graph presented in the experimental survey might not have been useful for most of the teams where a majority of the members had expertise under three years. One to three years of expertise is considered a beginner stage and not professional-level expertise. With this floor in expertise level, distinguishing between one or two years of experience likely did not mean a lot in terms of the expertise they could provide for collaboration.

None of the findings from previous studies on promoting team members' awareness on available expertise to enhance the development of TMSs in a laboratory setting were replicated in this study [17, 99, 109]. In one of the studies, participants who had been asked to evaluate other members in the same team divided their work more often than the participants who had not [17]. Another study showed that participants were able to identify the expertise of members more accurately in groups with high within-group expertise variability than the groups with low within-group variability [99]. The researchers manipulated the number of group members possessing knowledge of member expertise prior to actual collaboration as an experimental condition in order to see if the number had an effect on how teams would pool the necessary expertise for collaboration [109]. The study observed that more than half of members being aware of expertise was a prerequisite for them to exchange unique information for collaboration to occur. The failure to replicate the above results could be due to the messiness of real-world collaboration studied in this research.

8.5 Theoretical contributions

This research makes two theoretical contributions. First, this research extends the existing body of literature on TMS by examining the effect of an intervention designed to improve existing TMSs in teams through a field experiment. There is a paucity of literature on how to deliberately facilitate the development of TMS in real-world teams using any form of interventions. The majority of research on TMS has focused on inferring the effect of TMS in laboratory settings, or on retrospectively explaining the development of TMS in teams. As shown in the research model in Figure 8.1, this research accounted for a novel construct, the accurate awareness of available expertise in teams, as a potential lever to improve team coordination and performance. Although I was not able to demonstrate change in TMS due to adding this lever, further research may expand on this initial idea by conducting in-depth qualitative research on team coordination to identify another lever for cross-functional data science teams. Future research may also experiment with different ways of visualizing team work; for example, instead of revealing the misalignment in expertise perception for a short time period, researchers may test the effect of prolonged exposure to a visualization of team work on data science team members.

Second, to the line of research on understanding and supporting data science teams from the perspectives of Computer-supported Cooperative Work (CSCW), this research hypothesized and tested one of the interventions, highlighting inaccu-

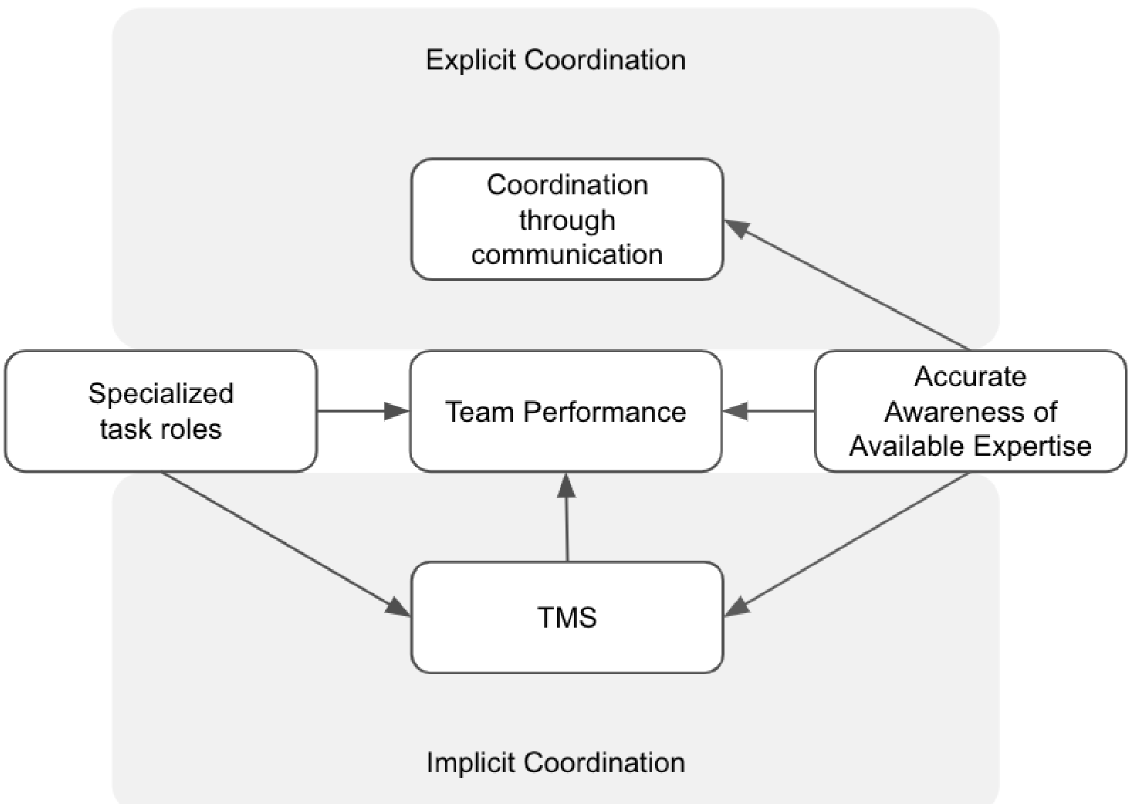

Figure 8.1: The conceptual research model of the dissertation

racies in the perceptions of available expertise in teams, which can be translated into the design of technology to support data science teams, for example in the design of user profiles for a collaborative platform for data scientists. Growing number of studies explored the ways to support data science teams by designing a novel collaborative platform for data scientists [110] or by investigating how individual data scientists exploited existing systems, not necessarily designed for data science collaboration, to support their collaborations [111]. The intervention tested in this research could be another strand of supportive measures for data science teams. Although this specific intervention was not conclusively demonstrated to affect TMS, coordination, or performance future work can expand on it by mining and visualizing the past work record, which could be more accurate representation of one's expertise than self-reports or aggregated accounts from team members. If a certain intervention that can be presented online can facilitate the development of TMSs in teams, the intervention could be used to enable coordination between members in remote data science teams.

8.6 Limitations and future research

Several limitations exist in this research. One, study participants in either Study 1 or 2 elected themselves to participate in the study so the sampling of participants was susceptible to self-selection bias. However, addressing and testing the same research questions in two of the same format of events a year apart could mitigate the self-selection bias by diversifying the samples and by testing if the same observations would uphold in the two samples. Two, the form of data science collaboration I studied in this research is rather specific with a relatively short collaboration period (i.e., a week), a competition setting, and mainly student participants. Due to this specificity, the findings may not generalize to the broader scope of data science collaborations, for example, the team work of data scientists in corporate settings. Yet the current setting of the research could shed a light on potential levers for building high-performing data science teams, a research subject with a growing interest.

Future work on improving team work in data science should uncover the effect of having members with specialized roles in teams on TMS in data science teams without previous collaboration experience. Dominant number of individuals who opted into this research, especially in Study 2, had known their team members or worked with them prior to the Data Challenge competition. This prior knowledge may have clouded the relationships between the constructs in the research model. If the information visualization used in this study could indeed facilitate the TMS in data science teams, it could be used to shorten the on-boarding time for newly formed data science teams, or to support virtual data science teams with no prior acquaintance between members.

Future work should identify objective measures of team performance for data science teams. In both Study 1 and 2, the judges failed to reach a reasonable level of consensus in rating the final submissions. In order for researchers to be able to evaluate and compare design implementations aimed at supporting data science teams, there needs to be objective measurements for team performance that quantify the effectiveness of the implementations.

8.7 Conclusion

In this study, I revealed how tasks were divided as a form of task roles among members of data science teams. I found that while technical roles were most often taken by team members (analyst and wrangler), non-technical roles were just as commonly assumed. Through a follow-up analysis of the identified roles, I discovered that assuming technical roles was about the presence of skills whereas taking on non-technical roles was about the absence of technical skills. Surprisingly, no strong connection was identified between a TMS and team performance measures, which is not consistent with previous studies. Covering more roles required for collaboration was argued to improve team performance, but it has not be replicated in the context of data science teams. This study is one of the few that cross-sectionally compared data science teams working on a similar format of problem within a relatively controlled setting in the format of competition.

Bibliography

[1] Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manage.*, 35(2):137–144, April 2015.

[2] Vasant Dhar. Data science and prediction. October 2012.

[3] Wil van der Aalst. Data science in action. In Wil van der Aalst, editor, *Process Mining: Data Science in Action*, pages 3–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[4] Ritu Agarwal and Vasant Dhar. Editorial—Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, 25(3):443–448, September 2014.

[5] Orhan Cam. 'data science is a team sport': DJ patil spends first day pitching silicon valley on joining government. `https://www.nextgov.com/analyticsta/2015/02/dataienceamorttilendsrstytchingliconlleyiningvernment/105648/`, February 2015. Accessed: 2019-4-22.

[6] Eric Colson. Why data science teams need generalists, not specialists. *Harvard Business Review*, March 2019.

[7] Daniel Tunkelang. Data scientists: Generalists or specialists? `https://www.oreilly.com/ideas/dataientistsneralistsecialists`, December 2015. Accessed: 2019-4-22.

[8] Paul R Carlile. A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science*, 13(4):442–455, 2002.

[9] T W Malone and K Crowston. What is coordination theory and how can it help design cooperative work systems? *Proceedings of the 1990 ACM conference on*, 1990.

[10] Thomas W Malone, Thomas W Malone, and Kevin Crowston. The interdisciplinary study of coordination. *ACM Comput. Surv.*, 26(1):87–119, March 1994.

[11] Gwen M Wittenbaum, Garold Stasser, and Carol J Merry. Tacit coordination in anticipation of small group task completion. *J. Exp. Soc. Psychol.*, 32(2):129–152, March 1996.

[12] Alberto Espinosa, F Javier Lerch, Robert E Kraut, Eduardo Salas, and Stephen M Fiore. Explicit vs. implicit coordination mechanisms and task dependencies: one size does not fit all. *Team cognition: understanding the factors that drive process and performance. American Psychological Association, Washington, DC*, pages 107–129, 2004.

[13] Elliot E Entin and Daniel Serfaty. Adaptive team coordination. *Hum. Factors*, 41(2):312–325, June 1999.

[14] J E Mathieu, T S Heffner, G F Goodwin, E Salas, and J A Cannon-Bowers. The influence of shared mental models on team process and performance. *J. Appl. Psychol.*, 85(2):273–283, April 2000.

[15] Daniel M Wegner, Toni Giuliano, and Paula T Hertel. Cognitive interdependence in close relationships. In William Ickes, editor, *Compatible and Incompatible Relationships*, pages 253–276. Springer New York, New York, NY, 1985.

[16] Kyle Lewis. Measuring transactive memory systems in the field: scale development and validation. *J. Appl. Psychol.*, 88(4):587–604, August 2003.

[17] Michael Robert Baumann. *The effects of manipulating salience of expertise and membership change on transactive memory.* PhD thesis, ProQuest Information & Learning, 2001.

[18] Nitesh Goyal, Gilly Leshed, Dan Cosley, and Susan R Fussell. Effects of implicit sharing in collaborative analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 129–138, New York, NY, USA, 2014. ACM.

[19] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. Strategies for crowdsourcing social data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 227–236, New York, NY, USA, 2012. ACM.

[20] Megan Smith. The white house names dr. DJ patil as the first U.S. chief data scientist. `https://obamawhitehouse.archives.gov/blog/2015/02/18/white-house-names-dr-dj-patil-first-us-chief-data-scientist`, February 2015. Accessed: 2019-5-8.

[21] Barack Obama. Transparency and open government. `https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government`, January 2009. Accessed: 2019-5-8.

[22] Brian Caffo, Roger D Peng, and Jeffrey Leek. Executive data science. `https://leanpub.com/eds`, December 2018. Accessed: 2019-5-8.

[23] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. The emerging role of data scientists on software development teams. In *Proceedings of the 38th International Conference on Software Engineering*, ICSE '16, pages 96–107, New York, NY, USA, 2016. ACM.

[24] Joohee Choi and Yla Tausczik. Characteristics of collaboration in the emerging practice of open data analysis. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 835–846, New York, NY, USA, 2017. ACM.

[25] Russell Jurney. Generalists dominate data science. `https://www.kdnuggets.com/2018/02/generalistsminatetaience.html`, February 2018. Accessed: 2019-5-8.

[26] Alain F Zuur, Elena N Ieno, and Chris S Elphick. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.*, 1(1):3–14, March 2010.

[27] G Andrienko, N Andrienko, and S Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations*, 2007.

[28] K Wongsuphasawat, D Moritz, A Anand, J Mackinlay, B Howe, and J Heer. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Trans. Vis. Comput. Graph.*, 22(1):649–658, January 2016.

[29] Nitesh Goyal and Susan R Fussell. Effects of sensemaking translucence on distributed collaborative analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 288–302, New York, NY, USA, 2016. ACM.

[30] Sharolyn Converse, J A Cannon-Bowers, and E Salas. Shared mental models in expert team decision making. *Individual and group decision making: Current issues*, 221, 1993.

[31] Richard Klimoski and Susan Mohammed. Team mental model: Construct or metaphor? *J. Manage.*, 20(2):403–437, April 1994.

[32] Laurie L Levesque, Jeanne M Wilson, and Douglas R Wholey. Cognitive divergence and shared mental models in software development project teams. *J. Organ. Behav.*, 22(2):135–144, March 2001.

[33] R R Patrashkova-Volzdoska, S A McComb, S G Green, and W D Compton. Examining a curvilinear relationship between communication frequency and team performance in cross-functional project teams. *IEEE Trans. Eng. Manage.*, 50(3):262–269, August 2003.

[34] Janis A Cannon-Bowers and Eduardo Salas. Reflections on shared cognition. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 22(2):195–202, 2001.

[35] Ralph H Turner. Role change. *Annu. Rev. Sociol.*, 16(1):87–110, 1990.

[36] R M Belbin. *Team Roles at Work*. Routledge, 2010.

[37] Robert B Cialdini, Carl A Kallgren, and Raymond R Reno. A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In Mark P Zanna, editor, *Advances in Experimental Social Psychology*, volume 24, pages 201–234. Academic Press, January 1991.

[38] Erving Goffman. *Encounters: Two Studies in the Sociology of Interaction*. Ravenio Books, 1961.

[39] Beth A Bechky. Gaffers, gofers, and grips: Role-Based coordination in temporary organizations. *Organization Science*, 17(1):3–21, February 2006.

[40] Melissa A Valentine and Amy C Edmondson. Team scaffolds: How mesolevel structures enable Role-Based coordination in temporary groups. *Organization Science*, 26(2):405–422, April 2015.

[41] Karl E Weick. The collapse of sensemaking in organizations: The mann gulch disaster. *Adm. Sci. Q.*, 38(4):628–652, 1993.

[42] R Meredith Belbin. Management teams: Why they succeed or fail. *Human Resource Management International Digest*, 2011.

[43] B Senior. Team roles and team performance: is there 'really'a link? *J. Occup. Organ. Psychol.*, 1997.

[44] Abhizna Butchibabu, Christopher Sparano-Huiban, Liz Sonenberg, and Julie Shah. Implicit coordination strategies for effective team communication. *Hum. Factors*, 58(4):595–610, June 2016.

[45] Harlan Harris, Sean Murphy, and Marck Vaisman. *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*. O'Reilly Media, Inc., 2013.

[46] M Kim, T Zimmermann, R DeLine, and A Begel. Data scientists in software teams: State of the art and challenges. *IEEE Trans. Software Eng.*, 44(11):1024–1038, November 2018.

[47] Jeffrey S Saltz and Nancy W Grady. The ambiguity of data science team roles and the need for a data science workforce framework. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 2355–2361. IEEE, 2017.

[48] Amy X Zhang, Michael Muller, and Dakuo Wang. How do data science workers collaborate? roles, workflows, and tools. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–23, 2020.

[49] Daniel M Wegner. Transactive memory: A contemporary analysis of the group mind. In Brian Mullen and George R Goethals, editors, *Theories of Group Behavior*, pages 185–208. Springer New York, New York, NY, 1987.

[50] Yuqing Ren and Linda Argote. Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *Ann. R. Coll. Physicians Surg. Can.*, 5(1):189–229, June 2011.

[51] Andrea B Hollingshead and Samuel N Fraidin. Gender stereotypes and assumptions about expertise in transactive memory. 2003.

[52] Julia Kotlarsky, Bart van den Hooff, and Leonie Houtman. Are we on the same page? knowledge boundaries and transactive memory system development in Cross-Functional teams. *Communic. Res.*, 42(3):319–344, April 2015.

[53] Jenny Liao, Anne T O'Brien, Nerina L Jimmieson, and Simon Lloyd D Restubog. Predicting transactive memory system in multidisciplinary teams: The interplay between team and professional identities. *J. Bus. Res.*, 68(5):965–977, May 2015.

[54] David P Brandon and Andrea B Hollingshead. Transactive memory systems in organizations: Matching tasks, expertise, and people. *Organization Science*, 15(6):633–644, December 2004.

[55] Zhi-Xue Zhang, Paul S Hempel, Yu-Lan Han, and Dean Tjosvold. Transactive memory system links work team characteristics and performance. *J. Appl. Psychol.*, 92(6):1722–1730, November 2007.

[56] Diane Wei Liang, Richard Moreland, and Linda Argote. Group versus individual training and group performance: The mediating role of transactive memory. *Pers. Soc. Psychol. Bull.*, 21(4):384–393, April 1995.

[57] Richard L Moreland and Larissa Myaskovsky. Exploring the performance benefits of group training: Transactive memory or improved communication? *Organ. Behav. Hum. Decis. Process.*, 82(1):117–133, May 2000.

[58] Kyle Lewis and Benjamin Herndon. Transactive memory systems: Current issues and future research directions. *Organization Science*, 22(5):1254–1265, October 2011.

[59] Ali E Akgün, John Byrne, Halit Keskin, Gary S Lynn, and Salih Z Imamoglu. Knowledge networks in new product development projects: A transactive memory perspective. *Information & Management*, 42(8):1105–1120, December 2005.

[60] Naina Gupta and Andrea B Hollingshead. Differentiated versus integrated transactive memory effectiveness: It depends on the task. *Group Dyn.*, 14(4):384–398, December 2010.

[61] John R Hollenbeck, Henry Moon, Aleksander P J Ellis, Bradley J West, Daniel R Ilgen, Lori Sheppard, Christopher O L H Porter, and John A Wagner, 3rd. Structural contingency theory and individual differences: examination of external and internal personam fit. *J. Appl. Psychol.*, 87(3):599–606, June 2002.

[62] Matthew J Pearsall, Aleksander P J Ellis, and Bradford S Bell. Building the infrastructure: the effects of role identification behaviors on team cognition development and performance. *J. Appl. Psychol.*, 95(1):192–200, January 2010.

[63] Betsy Sparrow, Jenny Liu, and Daniel M Wegner. Google effects on memory: cognitive consequences of having information at our fingertips. *Science*, 333(6043):776–778, August 2011.

[64] Sue Young Choi, Heeseok Lee, and Youngjin Yoo. The impact of information technology and transactive memory systems on knowledge sharing, application, and team performance: a field study. *Management Information Science Quarterly*, pages 855–870, 2010.

[65] Y Yoo and C Ifvarsson. Knowledge dynamics in organizations: Toward a theoretical framework. In D Willis E. Coakes and S Clarke, editors, *Knowledge Management in the Sociotechnical World: The Graffiti Continues*, pages 52–64. Springer-Verlag, London, 2001.

[66] John L Cordery and Christine Soo. Overcoming impediments to virtual team effectiveness. *Hum. Factors Man.*, 18(5):487–500, September 2008.

[67] Andrea B Hollingshead. Communication, learning, and retrieval in transactive memory systems. *J. Exp. Soc. Psychol.*, 34(5):423–442, September 1998.

[68] Terri L Griffith, John E Sawyer, and Margaret A Neale. Virtualness and knowledge in teams: Managing the love triangle of organizations, individuals, and information technology. *Management Information Systems Quarterly*, pages 265–287, 2003.

[69] D M Wegner, R Erber, and P Raymond. Transactive memory in close relationships. *J. Pers. Soc. Psychol.*, 61(6):923–929, December 1991.

[70] Garold Stasser, Dennis D Stewart, and Gwen M Wittenbaum. Expert roles and information exchange during discussion: The importance of knowing who knows what. *J. Exp. Soc. Psychol.*, 31(3):244–265, May 1995.

[71] Glenn E Littlepage, Andrea B Hollingshead, Laurie R Drake, and Anna M Littlepage. Transactive memory and performance in work groups: Specificity, communication, ability differences, and work allocation. *Group Dyn.*, 12(3):223–241, 2008.

[72] John R Austin. Transactive memory in organizational groups: the effects of content, consensus, specialization, and accuracy on group performance. *J. Appl. Psychol.*, 88(5):866–878, October 2003.

[73] Aleksander P J Ellis. System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. *AMJ*, 49(3):576–589, June 2006.

[74] Yuqing Ren, Kathleen M Carley, and Linda Argote. The contingent effects of transactive memory: When is it more beneficial to know what others know? *Manage. Sci.*, 52(5):671–682, May 2006.

[75] Ciaran Heavey and Zeki Simsek. Distributed cognition in top management teams and organizational ambidexterity: The influence of transactive memory systems. *J. Manage.*, 43(3):919–945, March 2017.

[76] Adrian Bangerter. Maintaining interpersonal continuity in groups: The role of collective memory processes in redistributing information. *Group Process. Intergroup Relat.*, 5(3):203–219, July 2002.

[77] Jessica Siegel Christian, Matthew J Pearsall, Michael S Christian, and Aleksander P J Ellis. Exploring the benefits and boundaries of transactive memory systems in adapting to team member loss. *Group Dyn.*, 18(1):69–86, March 2014.

[78] J Grudin. Computerpported cooperative work: history and focus. *Computer*, 27(5):19–26, May 1994.

[79] Joan Greenbaum and Morten Kyng. *Design at work: Cooperative design of computer systems*, volume 30. Lawrence Erlbaum Hillsdale, NJ, 1991.

[80] Petra Isenberg, Danyel Fisher, Sharoda A Paul, Meredith Ringel Morris, Kori Inkpen, and Mary Czerwinski. Cocated collaborative visual analytics around a tabletop display. *IEEE Trans. Vis. Comput. Graph.*, 18(5):689–702, May 2012.

[81] Aniket Kittur and Robert E Kraut. Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, pages 37–46, New York, NY, USA, 2008. ACM.

[82] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando P'erez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, and Others. Jupyter notebooks-a publishing format for reproducible computational workflows. In *ELPUB*, pages 87–90. books.google.com, 2016.

[83] Adam Rule, Ian Drosos, Aur'elien Tabard, and James D Hollan. Aiding collaborative reuse of computational notebooks with annotated cell folding. *Proc. ACM Hum. -Comput. Interact.*, 2(CSCW):150:1–150:12, November 2018.

[84] Matthew J Pearsall and Aleksander P J Ellis. The effects of critical team member assertiveness on team performance and satisfaction. *J. Manage.*, 32(4):575–594, August 2006.

[85] Richard L Moreland, Linda Argote, and Ranjani Krishnan. Socially shared cognition at work: Transactive memory and group performance. *What's social about social cognition? Research on socially shared cognition in small groups.*, 398:57–84, 1996.

[86] Li Lu, Y Connie Yuan, and Poppy Lauretta McLeod. Twenty-five years of hidden profiles in group decision making: a meta-analysis. *Pers. Soc. Psychol. Rev.*, 16(1):54–75, February 2012.

[87] Gerben S Van der Vegt, Ben J M Emans, and Evert Van De Vliert. Patterns of interdependence in work teams: a twovel investigation of the relations with job and team satisfaction. *Pers. Psychol.*, 54(1):51–69, 2001.

[88] Deborah L Gladstein. Groups in context: A model of task group effectiveness. *Adm. Sci. Q.*, 29(4):499–517, December 1984.

[89] Venugopal Balijepally, George Mangalaraj, and Kishen Iyengar. Are we wielding this hammer correctly? a reflective review of the application of cluster analysis in information systems research. *Journal of the Association for Information Systems*, 12(5):1, 2011.

[90] Joseph F Hair, Jr, William C Black, Barry J Babin, and Rolph E Anderson. *Multivariate Data Analysis (7th Edition)*. Pearson, 7 edition edition, February 2009.

[91] D L Davies and D W Bouldin. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1(2):224–227, February 1979.

[92] Ujjwal Maulik and Sanghamitra Bandyopadhyay. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on pattern analysis and machine intelligence*, 24(12):1650–1654, 2002.

[93] Ofer Arazy, Johannes Daxenberger, Hila Lifshitz-Assaf, Oded Nov, and Iryna Gurevych. Turbulent stability of emergent roles: The dualistic nature

of self-organizing knowledge coproduction. *Information Systems Research*, 27(4):792–812, 2016.

[94] Joachim Stempfle, Oliver Hübner, and Petra Badke-Schaub. A functional theory of task role distribution in work groups. *Group Process. Intergroup Relat.*, 4(2):138–159, April 2001.

[95] Georgine Beranek, Wolfgang Zuser, and Thomas Grechenig. Functional group roles in software engineering teams. In *Proceedings of the 2005 workshop on Human and social factors of software engineering*, HSSE '05, pages 1–7, New York, NY, USA, May 2005. Association for Computing Machinery.

[96] Linda Argote, Brandy L Aven, and Jonathan Kush. The effects of communication networks and turnover on transactive memory and group performance. *Organization Science*, 29(2):191–206, April 2018.

[97] Daniel G Bachrach, Kyle Lewis, Youngsang Kim, Pankaj C Patel, Michael C Campion, and Sherry M B Thatcher. Transactive memory systems in context: A meta-analytic examination of contextual factors in transactive memory systems development and team performance. *J. Appl. Psychol.*, 104(3):464–493, March 2019.

[98] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, et al. Jupyter notebooks-a publishing format for reproducible computational workflows. In *ELPUB*, pages 87–90, 2016.

[99] M R Baumann and B L Bonner. The effects of variability and expectations on utilization of member expertise and group performance. *Organ. Behav. Hum. Decis. Process.*, 2004.

[100] Leslie Gross Portney, Mary P Watkins, and Others. *Foundations of clinical research: applications to practice*, volume 892. Pearson/Prentice Hall Upper Saddle River, NJ, 2009.

[101] Ofer Arazy, Hila Liifshitz-Assaf, Oded Nov, Johannes Daxenberger, Martina Balestra, and Coye Cheshire. On the" how" and" why" of emergent role behaviors in wikipedia. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2039–2051. dl.acm.org, 2017.

[102] Diyi Yang, Aaron Halfaker, Robert Kraut, and Eduard Hovy. Who did what: Editor role identification in wikipedia. In *International Conference on Web and Social Media*, 2016.

[103] R Wageman, H Gardner, and others. The changing ecology of teams: New directions for teams research. *J. Organ. Behav.*, 2012.

[104] Heidi K Gardner, Francesca Gino, and Bradley R Staats. Dynamically integrating knowledge in teams: Transforming resources into performance. *AMJ*, 55(4):998–1022, August 2012.

[105] Ahsan Ali, Hongwei Wang, and Ali Nawaz Khan. Mechanism to enhance team creative performance through social media: A transactive memory system approach. *Comput. Human Behav.*, 91:115–126, February 2019.

[106] Julia Kotlarsky, Bart van den Hooff, and Marleen Huysman. Bridging knowledge boundaries in Cross-Functional groups: The role of a transactive memory system. In *ICIS 2009 PROCEEDINGS*, ICIS 2009 Proceedings, 2009.

[107] Julia Kotlarsky, Bart van den Hooff, and Leonie Houtman. Are we on the same page? knowledge boundaries and transactive memory system development in cross-functional teams. *Communic. Res.*, 42(3):319–344, April 2015.

[108] Jenny Liao, Anne T O'Brien, Nerina L Jimmieson, and Simon Lloyd D Restubog. Predicting transactive memory system in multidisciplinary teams: The interplay between team and professional identities. *J. Bus. Res.*, 68(5):965–977, May 2015.

[109] M R Baumann and B L Bonner. Member awareness of expertise, information sharing, information weighting, and group decision making. *Small Group Research*, 2013.

[110] Kevin Crowston, Jeff S Saltz, Amira Rezgui, Yatish Hegde, and Sangseok You. Socio-technical affordances for stigmergic coordination implemented in MIDST, a tool for Data-Science teams. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW):1–25, November 2019.

[111] Ruqin Ren, Bei Yan, and Lian Jian. Show me your expertise before teaming up: Sharing online profiles predicts success in open innovation. *Internet Research*, ahead-of-print(ahead-of-print), January 2019.