Diversity and team performance: A series of field experiments
Hoogendoorn, S.M.

Citation for published version (APA):
Hoogendoorn, S. M. (2013). Diversity and team performance: A series of field experiments
Chapter 1

Introduction

1.1 Motivation

This dissertation reports about a series of field experiments conducted to measure the impact of diversity in teams on their performance. Over the past decades, teams rather than individuals have become increasingly relevant for organizational decision making and performance (Hamilton et al., 2003; Woolley et al., 2010). As Hamilton et al. (2003) put it: “during the past 30 years the use of teams has become a mainstay for the organization of work” (p. 465). Examples include various sorts of organizations such as government bodies, judges in collegial courts, academic researchers, boards of directors, business units, project teams, but also most venture start-ups are nowadays founded by teams (Parker, 2009). As a consequence, the effective composition of teams has emerged as a valuable asset for organizations and a promising field of research for academics (Hamilton et al., 2012). Arguably, one of the most influential determinants of the effectiveness of teams is their diversity (Page, 2007).

Diversity is commonly associated with benefits in terms of a more diverse pool of talents, improvement of an organization’s identity or brand image, better stakeholder relationships, increased employee motivation and higher customer satisfaction (e.g., Desvaux et al., 2007). As such, diversity can be employed as an organizational tool to promote efficiency or profitability and has become a momentous topic among management practitioners, policy makers and interest groups. Other arguments that are frequently used in the discussion about diversity concern inclusion of (minority) groups in society, fairness considerations and attenuation of stereotypes (due to exposure to dissentient individuals). Moreover, it is often related to benefits such as complementarities and (mutual) learning that could become apparent in more creativity and better decision making. However, diversity may also coincide with higher costs of communication and coordination that possibly manifest themselves in more conflicts and a worse atmosphere (Lazear, 1999).
It is *a priori* ambiguous whether the costs or benefits of diversity prevail in different contexts and, hence, the net effect of diversity on team performance remains a question that needs to be answered empirically. Indeed, the relationship between team performance and diversity has been extensively studied: (i) in various strands of literature such as economics and finance, entrepreneurship, organization and management, psychology and sociology; (ii) along a number of dimensions such as (cognitive) ability, age, ethnicity, experience, gender, norms and personality; (iii) considering different units of analysis such as entire organizations, teams in the workplace, start-up and top management teams (see Shore et al., 2009).1 The vast majority of studies in this field of research, however, measures the *link* between diversity and performance of teams rather than the *effect* of diversity on team performance. Team composition in organizations is not rarely driven by productivity expectations of managers or self-selection of employees into and out of teams (see example below). Hence, the measured effects of diversity on performance in real-world teams are likely to be biased due to endogenous team composition (Adams et al., 2010; Hermalin and Weisbach, 2003).

Most studies in this field of research also measure several dimensions of diversity simultaneously, which prevents investigation of each diversity dimension in isolation. This is particularly relevant when interested in the bare effect of a specific diversity dimension on team performance. For example, what is the bare impact of gender diversity on the performance of teams without measuring the confounding effects of diversity in age or ethnicity at the same time? The objective of this dissertation is to disentangle the impact on performance of various dimensions of diversity in teams, one by one. To this end, three field experiments were conducted with a focus on the separate effects of diversity in gender, ethnicity and cognitive ability. Such evidence is potentially useful for the optimal composition of teams in public and private organizations.

The relevance of measuring the *effect* of diversity on team performance could perhaps be illustrated by the current debate on more gender diversity in boards of directors. Arguments in favor of more gender diversity are often based on studies that show a positive relationship between the share of women in corporate boards and firm performance. However, this positive link does not allow for the interpretation that gender diversity in boards of directors *causes* firms to perform better. Women may self-select into high performing firms or, as a (probably not very well informed) devil’s advocate would possibly argue, high performing firms can afford to have non-value enhancing women on their boards of directors. In sum, studies into the business case for diversity on the basis of observational data are conceivably plagued by various endogeneity issues.

---

1A more detailed review of the literature on the relationship between different types of diversity, team performance and underlying mechanisms is provided in the introductions of the core chapters.
Experiments in the laboratory typically do not suffer from endogenous team composition and have yielded results that are largely consistent with the trade-off between the benefits and costs of more diversity (as discussed above). The resemblance of laboratory experiments to real-world situations, however, may be limited. This especially holds for the topic at hand: the effects of team diversity on complementarities, (mutual) learning, communication and coordination are not likely to become evident instantaneously (Boisjoly et al., 2006). It usually takes time for teams to establish roles, build up relationships, and create routines and processes to achieve their common goal. Studying the causal impact of diversity in teams on their longer run performance in realistic though controlled circumstances appears to be a challenging task. Consequently, studies that pursue this kind of measurement are scarce. A method that potentially shares the strong features (or avoids the weaknesses) of studies based on observational data and laboratory experiments is that of a field experiment (Bardsley et al., 2010; Harrison and List, 2004).

1.2 Method and contribution

This dissertation measures the impact of diversity on team performance in the longer run using field experiments among teams of undergraduate students that start up, manage and liquidate a real company as a compulsory and significant part of the first-year curriculum of an international business program in the Netherlands. It exploits the unique opportunity that these student companies are simultaneously founded and dissolved (after one academic year) in a controlled though rather realistic setting and that team composition in terms of gender, ethnicity and cognitive ability could be exogenously varied.

Students in the entrepreneurship program are required to select a business activity, raise capital by issuing shares, appoint officers and delegate tasks, produce and market products or services, keep the accounts and conduct shareholders’ meetings. Everything about the company is real (including tax and social security payments) and students face strong incentives that align their interests with the business outcomes of the company. Hence, teams that consist of approximately 10 to 12 students on average execute a substantial and genuinely joint task in order to maximize sales, profits and shareholder value. The performance of teams throughout this dissertation is measured in terms of these business outcomes. Additional questionnaires that were administered before, halfway and after the program enable to explore various underlying mechanisms suggested in the literature that possibly explain the effect of diversity on team performance such as complementarities, (mutual) learning, communication and coordination (Lazear, 1999).
The three field experiments were conducted in the academic years of 2008-2009 and 2009-2010. In the first academic year, 550 students were randomized into 45 teams conditional on their gender and ethnicity to construct exogenous variation in the ethnic composition and gender composition of teams. Since gender diversity and ethnic diversity are orthogonal in the sample of teams this will not compromise the design of both field experiments. In the academic year of 2009-2010, the ability composition of teams was manipulated by randomly assigning 573 students to 49 teams given their measured cognitive abilities. Both students and professors in these field experiments were kept uninformed about the stratified randomization in gender, ethnicity and cognitive ability. To avoid repetition, the descriptions of the field experiments (context, design and incentives) are combined into a separate chapter (Chapter 2). The empirical and statistical methods used are described in the core Chapters 3 through 5 that discuss the field experiments on the separate impact of diversity in gender, ethnicity and cognitive ability, respectively. For various reasons these methods are not always uniform across the different core chapters.

The field experimental design that is employed in this dissertation addresses concerns regarding self-selection, omitted variable bias and reversed causality in a setting that more closely resembles the functioning of teams than previous laboratory studies. Moreover, the experimental set-up allows to construct a large variation in diversity between teams, relative to studies based on observational data, which is of interest for examining possible non-linearities in the relationship between team performance and diversity. Finally, the three field experiments facilitate separate and unconfounded inference of the effects of diversity in gender, ethnicity and cognitive ability on the performance of teams. This dissertation takes advantage of the novel opportunity to conduct field experiments with stratified randomization to measure the effects of diversity in teams.

The internal validity of this experimental approach comes at a cost. It exploits exogenous variation among students who jointly run a company for a year instead of individuals working together in teams for a longer period with a possibly infinite time horizon. Students in these field experiments are roughly from the same caliber, but lack serious work experience which potentially limits the generalizability of the results (to teams in public and private organizations). A second limitation of the experimental method is that it cannot measure effectively what kinds of underlying mechanisms explain possible differences in performance between teams. Teams of high diversity, for example, may be characterized by more complementarities and (mutual) learning. However, this does not exclude that team performance has a reverse effect on these mechanisms, i.e., the mechanisms of interest are endogenous. As a consequence, this dissertation (like other studies) provides at best suggestions why teams of distinct degrees of diversity may perform differently.
Despite the first limitation, the experimental design used has some external validity. Students execute a substantial and genuinely joint task that requires them to establish roles, build up relationships, and create routines and processes in a similar way as managers or employees proceed in teams. Real-world examples of teams with similarly broad and complex tasks are internal business, start-up and top management teams. All in all, it is likely that the field experiments presented in this dissertation closely resemble the functioning of many sorts of teams in practice and, hence, potentially improve our understanding of the impact of diversity in real-world teams on their performance.

1.3 Outline

Chapter 2 discusses the context and design of the field experiments as well as the incentives that members experience to align their interests with those of the team.

Chapter 3 examines the effect of gender diversity on the performance of teams. It is based on the paper 'The impact of gender diversity on the performance of business teams: Evidence from a field experiment', co-authored with Hessel Oosterbeek and Mirjam van Praag (forthcoming in Management Science, 2013). As mentioned above, this chapter may shed light on the current discussion about the underrepresentation of women in boards of directors. In 2010, women held just above 15 percent of board seats at Fortune 500 companies, only 10 percent of board seats at the top 300 European companies and even less in Dutch listed companies (Catalyst, 2010; Woods, 2010).

From a labor supply perspective, increases in the share of women in higher education and the labor force have strengthened the policy relevance of more gender diversity in business (management) teams. Some countries including the Netherlands even enforce or are planning to enforce a higher representation of women by implementing minimum quota of board seats for female directors. So far there was little evidence that these regulations were in line with better firm or team performance. The results from the field experiment in this chapter show that gender diverse teams perform better than male-dominated teams. In addition, the results suggest that gender diverse teams perform no worse than teams with a majority of females. The data do not support any of the various underlying mechanisms that are tested to possibly explain why teams with an equal gender mix perform better.

Chapter 4 studies the impact of heterogeneity in ethnicity on team performance. It is based on the paper 'Ethnic diversity and team performance: A field experiment', co-authored with Mirjam van Praag. Ethnic diversity is highly relevant in a globalized world, the share of ethnic minorities in Western populations is increasing sharply and it is therefore likely that any team will become more diverse in terms of ethnicity (Alesina
and La Ferrara, 2005). For example, in the United States the share of minorities is expected to rise from about one-third nowadays to roughly the majority in 2042 (Bernstein and Edwards, 2008).

The field experiment in this chapter might provide a realistic preview of the effect that a high degree of ethnic diversity may have on the performance of teams. The results indicate that a moderate level of ethnic diversity has no impact on team performance, but if at least the majority of team members is ethnically diverse then more ethnic diversity has a positive effect on team performance. In line with theoretical predictions, this positive effect could be related to the more diverse pool of relevant knowledge facilitating (mutual) learning within ethnically diverse teams.

Chapter 5 investigates whether teams’ diversity in cognitive ability affects their performance. It is based on the paper ‘Ability dispersion and team performance: A field experiment’, co-authored with Simon Parker and Mirjam van Praag. Cognitive ability shapes individuals’ behavior, decision-making and performance outcomes (Cutler and Lleras-Muney, 2010; Dohmen et al., 2010; Grinblatt et al., 2011), and is a major determinant of individual earnings, income distribution and - at the aggregate level - economic growth (Hanushek and Woessmann, 2008). Yet, despite it being widely believed that (cognitive) abilities of members affect overall team performance, the precise impact of ability dispersion on the performance of teams remains poorly understood (Hamilton et al., 2012).

This chapter proposes a model in which greater ability dispersion generates greater knowledge for a team, but also reduces the effectiveness of monitoring to detect (and punish) shirking. Consistent with the predictions of the model and controlling for average ability, team performance first increases and then decreases with ability dispersion. Teams of moderate ability dispersion also experience fewer dismissals due to a lower degree of shirking members, although this does not chiefly explain why those teams achieve better results.

Chapter 6 briefly summarizes the findings from the previous chapters and provides some concluding remarks, including possible ideas for future research.