Diversity and team performance: A series of field experiments

Hoogendoorn, S.M.

Publication date
2013
Document Version
Final published version

tinbergen institute

Diversity and Team Performance

A Series of Field Experiments

Sander Hoogendoorn

Universiteit van Amsterdam

DIVERSITY AND TEAM PERFORMANCE: A SERIES OF FIELD EXPERIMENTS

ISBN 978 90 361 0348 0

Cover design: Crasborn Graphic Designers bno, Valkenburg a/d Geul

This book is no. 555 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

DIVERSITY AND TEAM PERFORMANCE: A SERIES OF FIELD EXPERIMENTS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 8 maart 2013, te 16:00 uur

door

Sander Matthijs Hoogendoorn

geboren te Harderwijk

PROMOTIECOMMISSIE

Promotor:

Prof. dr. C.M. van Praag

Overige leden:

Prof. dr. R.W. Fairlie
Dr. P. Koellinger
Prof. dr. T.J.S. Offerman
Prof. dr. R. Sloof

Faculteit Economie en Bedrijfskunde

Acknowledgements

This dissertation marks the end of an intense but rewarding period of my life. I would like to seize the opportunity to thank several people for their great help, one way or the other. First and foremost, I am grateful to my supervisor Mirjam van Praag who slapped my fingers if necessary (figuratively speaking) and encouraged me whenever possible. Her feedback, guidance and support were indispensable. I admit that it may sound somewhat clichéd (though from the heart), but I probably could not have wished for a better supervisor. My gratitude also goes to Hessel Oosterbeek and Simon Parker who I have come to see as my informal supervisors. Their comments, suggestions and contributions improved this dissertation tremendously. Furthermore, I am indebted to my girlfriend Mariëlle van Reisen who tolerated my sometimes intolerable behavior, gave me the opportunity to decompress from work and, hence, shared her infinite love and understanding.

I would like to thank my colleagues for providing such a vibrant and stimulating work environment. I really enjoyed our discussions about research, teaching and less related (but often hilarious) topics, particularly with my upstairs neighbor Adam Booij. This dissertation benefited enormously from their criticism, help and friendship. I also greatly appreciate Rob Fairlie, Philipp Koellinger, Theo Offerman and Randolph Sloof for their willingness to be part of my doctorate committee, and Silvia Dominguez Martinez and Laura Rosendahl Huber who were so kind to act as my paranymphs. Moreover, (financial) support from the Dutch Ministry of Economic Affairs and the Department of International Business Studies of the Amsterdam College of Applied Sciences is gratefully acknowledged. Last but certainly not least, I am very thankful to my family and friends for their patience, enthusiasm and company throughout the dissertation process. Thank you all so much!

Sander Hoogendoorn

Amsterdam, December 2012

Contents

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.

[1] A 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.

Chapter 2

Context, design and incentives of the field experiments

This chapter describes the context and design of the field experiments, discussed in Chapters 3 through 5 of this dissertation, as well as the incentives that members experience to align their interests with those of the team.

2.1 Context

Teams in all three field experiments take part in the Junior Achievement (JA) Young Enterprise Start Up Program, in the US known as the JA Company Program. This is the worldwide leading entrepreneurship program in secondary education (US and Europe) and post-secondary education (only Europe). The number of students participating in these JA-programs is substantial and steadily rises over the years (see Oosterbeek et al., 2010).

The entrepreneurship program is a compulsory part of the curriculum at the department of international business studies of the Amsterdam College of Applied Sciences, where the three field experiments of this dissertation take place. It lasts for an entire academic year and covers about one-fifth of students' first-year undergraduate curriculum in all sub-departments/fields of study.[1] The program is not a business simulation and requires students to start up, manage and liquidate a real company with a team of about 10 to 12 fellow students on average. Each company has to report to their randomly assigned professor and business coach on a regular basis.

[1] The international business program at the Amsterdam College of Applied Sciences is divided into five sub-departments/fields of study: business management, management, trade management Asia, business languages and financial management. This division is relevant since the assignment of students to teams takes place within these sub-departments/fields of study (discussed in Section 2.2).

Companies typically operate as follows. They start with brainstorming about potential business activities and conducting market research to select the most viable idea. There are no restrictions on the type of business activity that can be chosen. Simultaneously, teams appoint about half of their members to management positions (including the CEO and CFO) and the other half to non-management positions. Management positions are redistributed among the non-managing part of the team halfway the program. Management teams are a representative part of the entire team in terms of their diversity (see Section 2.2).

Subsequently, companies further develop their chosen idea by writing a business plan, and they start raising capital by issuing shares. Other sources of financing such as personal or outside loans are prohibited. Once the business plan is authorized by the majority of shareholders at the first shareholders' meeting, the key activities of teams boil down to production and marketing of the chosen products or services. All companies are dissolved at the end of the program and each team has to write an annual report that needs approval at the final shareholders' meeting. Any profits are proportionally divided among the shareholders. In sum, teams perform a substantial and genuinely joint task in which they have to establish roles, build up relationships, and create routines and processes to achieve their common goal.

2.2 Design

The three field experiments reported in this dissertation were performed in the academic years of 2008-2009 and 2009-2010. In the first academic year exogenous variation was constructed in the gender composition and ethnic composition of teams, while in the academic year of 2009-2010 the ability composition of teams was experimentally manipulated.[2] In all three field experiments students were randomized into teams conditional on their gender, ethnicity or cognitive ability one week before the start of the entrepreneurship program (when administrative data were received to assist in assigning students to teams). In practice, the stratified randomization proceeded as follows.

To construct differences in gender diversity across teams the fraction of female students was determined and varied for each team within fields of study. Male and female students were then randomly assigned to teams in accordance with these fractions. In a similar fashion, exogenous variation was imposed with regard to the ethnic composition of teams. Students are defined as native or Dutch if both parents are born in the Netherlands, whereas students are considered immigrant, minority or non-Dutch if at

[2]Since ethnic diversity and gender diversity are orthogonal this will not affect the results of both field experiments. The correlation is -0.185 and not significantly different from zero.

least one parent is not born in the Netherlands (see Hartog, 2011). The ethnicity of students was determined based on students' names, Dutch versus non-Dutch (see Mateos, 2007 for a motivation of this name-based procedure). Nearly 90 percent of these name-based classifications matched with students' actual ethnicity that was retrieved through the pretreatment survey one week later. As expected, most of the mismatches in the category of ethnic minority students occured among those with Dutch fathers. Table 2.1 lists team characteristics such as their assigned gender and ethnic composition as well as size, business performance and the product or service sold in the academic year of 2008-2009.

In the academic year of 2009-2010, the ability composition of teams was manipulated by randomly assigning students to teams conditional on their score in the 20-minute timed version of Raven's advanced progressive matrices test as a proxy for cognitive ability (Hamel and Schmittmann, 2006; Raven et al., 1998).[3] Within fields of study, students were divided into four quartiles per class on the basis of their test score, where 1 reflects the best quartile and 4 the worst quartile (classes were randomly composed within fields of study). Each class was then split up in two teams, which received either treatment A or B. Treatment A combines cognitive ability quartiles 1+2 and 3+4, and treatment B combines cognitive ability quartiles 1+4 and 2+3 in a class. Hence, '1+2 teams' have a high average ability and a low ability dispersion, '3+4 teams' have a low average ability and a low ability dispersion, '1+4 teams' have a medium average ability and a high ability dispersion, and '2+3 teams' have a medium average ability and a low ability dispersion. Table 2.2 reports the key characteristics of all teams.

The assignment of students to teams was enforced by the program coordinators who were informed about the nature of the stratified randomization. A few late applicants were randomly distributed among the existing teams, whereas a few 'no shows' were also randomly distributed across teams (as they did not know to which team they were assigned to at that stage). Students and business coaches were uninformed, while professors only knew that a research project was conducted which required to stick to the imposed team assignment. In both academic years only 6 students managed to switch teams (out of respectively 550 and 573 students). Differences with assignment procedures in previous years were kept to a minimum. Like in previous years team composition is not self-selected by students, but enforced by the college. Interviews with students corroborate their ignorance regarding the stratified randomization.

[3]Students were kept uninformed about their score in the 20-minute timed version of Raven's advanced progressive matrices test. The fact that they were tested was presented to them as a standard procedure of the introductory week at their new college.

Table 2.1. Team characteristics in the academic year of 2008-2009

#	Name	Share of women	Share of minorities	Team size	Sales (euros)	Profits (euros)	Profits/ share	Description of product/service
1	A-Card	0.25	0.50	16	1236	-848	-11.78	Discount card Amsterdam nightlife
2	A'dam Gadgets	0.36	0.45	12	534	-41	-0.47	USB hot plate for coffee, tea, etc.
3	Appie	0.90	0.40	11	455	150	3.00	Apple-shaped box to preserve apples
4	Aqua de Coctail	0.42	0.58	12	1130	-306	-3.12	Comprehensive cocktail shaker set
5	ArtEco Bags	0.40	0.60	11	912	-402	-7.44	Durable give-away bag clothes stores
6	BubbleMania	0.18	0.70	11	503	-62	-1.34	Multi-purpose protective key chain
7	D'Wine	0.25	0.63	9	740	-55	-1.62	Bottles of wine
8	Eastern Green	0.36	0.69	14	513	106	2.93	Engravable text bean growing a plant
9	Escapade Inc	0.67	0.22	9	593	-111	-3.09	Tube clip for sealing food, toiletry, etc.
10	eyeBMA	0.38	0.50	16	558	125	3.90	Package with easy-to-use eye shadow
11	Firefly	0.50	0.20	12	2226	294	3.67	Ascending fire lantern for celebrations
12	Fl!pthat	0.23	0.64	13	455	215	9.77	Redecorating already existing websites
13	Ginger	0.58	0.58	12	977	-107	-2.14	Multi-purpose solar energy charger
14	Himitsu	0.30	0.86	10	775	36	0.86	n/a
15	I-Care	0.38	0.54	15	1204	477	11.36	Dead Sea minerals beauty products
16	iJoy	0.36	0.64	14	1953	94	1.44	Wristband with USB capacity
17	I-Juice	0.38	0.54	13	1255	-39	-0.42	Pocket-size lightweight mobile charger
18	IMSC	0.27	0.55	11	625	-390	-7.41	n/a
19	iShield	0.44	0.50	11	4209	130	2.20	Invisible protective shield for iPhones
20	KISBag	1.00	0.57	9	205	-117	-3.90	Tiny foldable bag replacing plastic bags
21	Laservibes	0.36	0.40	11	130	-229	-4.32	Organizing lasershows for companies
22	Mengelmoes	0.33	0.71	10	942	63	1.24	Easy-to-wear telephone charger device
23	My-Buddy	0.17	0.45	12	297	-58	-2.65	USB doll for kids reflecting emoticons
24	Nine2Five	0.73	0.60	12	235	-1016	-12.87	USB hot plate for coffee, tea, etc.
25	Picture Perfect	0.21	0.54	15	260	-51	-1.45	Customized shirts for men and women

Table 2.1. (continued)

#	Name	Share of women	Share of minorities	Team size	Sales (euros)	Profits (euros)	Profits/ share	Description of product/service
26	Pietje Plu	0.73	0.40	12	n/a	n/a	n/a	Trendy umbrellas
27	Pocket Memory	0.38	0.73	16	979	103	1.20	Business cards with USB capacity
28	Pro'Lux	0.31	0.54	14	378	-395	-9.18	Promotional gifts with USB capacity
29	Qwinlok	0.31	0.42	13	340	35	0.91	Boxer shorts for female adolescents
30	Reflection	0.82	0.36	11	890	45	0.84	Cosmetics mirror incl. mascara clip
31	SAME	0.82	0.36	11	1618	152	2.15	Comfortable unisex earwarmer
32	Sappho	0.50	0.50	8	980	n/a	n/a	n/a
33	Sharity	0.58	0.67	12	265	-241	-8.04	Peace sign necklace for teenagers
34	ShoeTattoo	0.62	0.77	13	270	88	1.21	Shoe customization by graphic artists
35	Student Promotion	0.42	0.42	13	571	235	15.64	Promotional activities for companies
36	StuPill	0.38	0.31	14	731	-1011	-15.48	Convenient Indonesian anti-RSI pillow
37	Test-a-Holic	0.45	0.45	11	728	220	4.88	Alcohol breath tester for nightlife
38	We-Do Solutions	0.10	0.56	10	604	-267	-6.06	Multi-purpose trendy key chain
39	We 'R U	0.33	0.33	13	1041	50	0.89	Compact wallet in several colors
40	XNG	0.50	0.90	12	1088	258	7.60	Shirts "Chicks on Kicks" community
41	YEN Empowered	0.50	0.83	13	1267	33	0.71	n/a
42	YET's Wear	0.53	0.79	16	789	-247	-2.47	Customized shirts own YET-brand
43	YOU	0.17	0.64	12	0	-242	-6.55	Hotel door hanger to store keys, etc.
44	Young Legends	0.44	0.67	9	400	59	0.84	n/a
45	YUVA	0.53	0.70	16	1153	294	12.79	Engravable rice grain in glass covering

Note: Share of women is based on a dummy indicator for male and female students, whereas share of minorities is based on a dummy indicator for students of Dutch and non-Dutch ethnicity (excluding students whose gender or ethnicity is unknown). The number of students whose gender is unknown amounts to 20 and the number of students whose ethnicity is unknown adds up to 52 (out of 550 students). Team size reflects the size of teams at baseline. Missing or incomplete descriptions of a team's product or service are indicated by 'n/a' (not available).

Table 2.2. Team characteristics in the academic year of 2009-2010

#	Name	Average ability	CV ability	Team size	Sales (euros)	Profits (euros)	Profits/ share	Description of product/service
1	Achter de rits	17.64	0.33	11	534	78	1.56	Sustainable fair trade boxer short
2	Be Portable	16.00	0.22	10	1180	288	4.65	Portable durable water bottle
3	Berry Emotions	21.45	0.26	12	88	-592	-13.76	Personalized telephone cover
4	Blue Empire	14.09	0.19	13	418	-75	-1.80	Fashionable bluetooth bracelet
5	Cleaning Env. Friendly	14.75	0.31	8	416	4	0.06	Long-lasting durable paper towel
6	Chori-Chori	16.62	0.20	15	588	-441	-3.94	Trendy customized Indian jewelry
7	Comb. Young E'ship	15.13	0.19	9	894	120	2.45	Whistling operated key finder
8	Covers 'n Cases	19.36	0.37	16	339	56	1.52	Covers and cases for telephones
9	Easylife	19.13	0.32	12	207	-289	-8.25	Automated toothpaste dispenser
10	Flashing Productions	16.70	0.31	12	616	90	1.96	Fluorescent nightlife bracelet
11	FresHES	15.08	0.20	15	3505	146	3.10	Wine bottle shaped umbrella
12	Freshual	22.08	0.15	13	1600	353	4.58	Convenient compact key torch
13	Green	19.50	0.21	10	524	112	4.66	Unbreakable vase for flowers
14	Green Solutions	14.07	0.31	14	1576	376	9.89	Environmental awareness products
15	Horanje	17.38	0.33	14	464	37	0.52	Gift bag with Dutch souvenirs
16	Ibrella	17.18	0.18	12	239	-474	-6.97	Easy-to-wear mini umbrella
17	I-corporation	16.27	0.22	15	43	-637	-9.65	Convenient eye-drop dispenser
18	I'Magine	18.83	0.21	15	3268	703	5.81	Fashionable Brazilian flip-flops
19	Innovart	22.40	0.16	12	42	-8	-0.20	Broker/intermediary in art
20	Jikoni	18.56	0.20	10	861	54	1.36	Innovative kitchen appliances
21	Kamikare	17.50	0.18	10	578	27	0.52	Easy-to-use hair styling product
22	K'deau	19.83	0.32	17	878	120	1.74	Gift box with various gadgets
23	Keyen	19.00	0.13	13	1526	256	3.41	High-quality key finder device
24	Lightdancer	16.17	0.47	13	140	-596	-7.84	Jewelry with built-in laser beam
25	MagicTouch	21.86	0.08	7	540	-533	-5.49	Compact women-friendly vibrator

Table 2.2. (continued)

#	Name	Average ability	CV ability	Team size	Sales (euros)	Profits (euros)	Profits/ share	Description of product/service
26	Muffle	17.78	0.36	11	1321	331	6.13	Trendy coffee to-go cup sleeve
27	Nailtastic	17.78	0.15	10	285	-285	-7.49	Nailpolish applicator help kit
28	New Young Pros	20.43	0.30	10	1476	794	12.41	Anti-slip mat for multiple devices
29	Razor2Go	18.42	0.08	13	388	-113	-1.83	Comfortable facial hair remover
30	ReWrite	14.75	0.25	11	1600	413	6.36	T-shirt with rewritable tekst area
31	Rising Sun	14.78	0.30	11	578	-113	-4.71	Mini high-definition pen camera
32	Sleevies	20.13	0.25	10	5049	1619	25.70	Public transport chip card holder
33	Stud. Offer Products	22.22	0.18	9	144	-218	-4.35	Aids awareness-raising key chain
34	Sunny & Co	14.00	0.25	10	653	195	4.23	Wristband with sunscreen notifier
35	SunVision	20.40	0.17	10	392	-197	-5.81	Affordable sunglasses on strength
36	Swops	23.22	0.12	9	557	113	1.28	Trendy digital slim sports watch
37	Cardholder Company	18.11	0.12	9	252	-140	-5.60	Cardholder wallet with mirror
38	The Mug	20.70	0.15	13	2019	508	5.98	Comfortable coffee to-go mug
39	TieSupply	20.38	0.07	10	434	-243	-1.80	Tie roll cylinder in a travel case
40	Toccare	21.13	0.12	15	464	-226	-5.65	Key chain with digital business card
41	Winterwarmth	20.11	0.21	13	893	481	14.56	Multi-purpose hot and cold gel pad
42	Wristbands4World	17.69	0.14	14	596	51	1.64	Environmental awareness wristband
43	Xihe	20.70	0.09	11	861	-993	-10.13	Innovative solar battery charger
44	XXL	21.08	0.19	12	2553	239	2.91	HES University hooded sweatshirt
45	BubbleBeat	17.75	0.35	8	0	-225	-3.88	n/a
46	Cotton&Class	21.25	0.30	10	1201	-352	-4.69	n/a
47	Decent	19.25	0.08	13	210	-2	-0.08	n/a
48	InOffice	21.11	0.15	10	475	154	7.33	n/a
49	WatchUs	21.75	0.10	13	717	193	8.02	n/a

Note: Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test (excluding students whose test score is unknown). The number of students whose test score is unknown amounts to 60 (out of 573 students). Team size reflects the size of teams at baseline. Missing or incomplete descriptions of a team's product or service are indicated by 'n/a' (not available).

2.2.1 Composition dynamics

Composition dynamics that take place after the stratified randomization of students into teams may compromise the design of the three field experiments if these dynamics alter the imposed exogenous variation in terms of gender, ethnicity or cognitive ability. This subsection discusses to what extent dropouts during the entrepreneurship program and the appointment of members to management teams (possibly the more influential part of the team) changed the initial team composition.

Dropout rates for first-year students in Dutch higher vocational schools are high at about 30% including students that switch study and/or school (ref. HBO-raad, 2010). A possible explanation might be that the admission of students based on grades or previous achievements is not allowed in the Netherlands. The number of dropouts was equal to 104 students (19%) during the field experiments that were conducted in the academic year of 2008-2009, while 82 students (14%) dropped out during the field experiment that was performed in the academic year of 2009-2010.[4] This reduced the average team size in both academic years from approximately 12 at the start to about 10 at the end of the program.

The design of the field experiment on the impact of gender diversity was not contaminated by dropouts. Dropouts neither affected a team's gender composition nor varied across teams in relation to their gender composition. The regression coefficient between teams' share of women at baseline and at the end of the program is not significantly different from 1 and regressing students' dropout status on the share of women at baseline returns an insignificant point estimate. Similarly uncontaminating dropout dynamics occurred in the field experiments on the impact of diversity in ethnicity and cognitive ability.

Another composition dynamic that takes place after the initial assignment of students to teams is that companies appoint a management team twice (for the first and second half of the program). The composition of the entire team - on which the different treatments are based - may not be a realistic reflection of the management team. F-tests indicate that the composition of the management team in terms of gender, ethnicity or cognitive ability closely resembles that of the entire team. The coefficients from a regression of the gender, ethnic or ability composition of the management team on the gender, ethnic or ability composition of the entire team are not significantly different from 1, neither before nor after the roles of manager and employee have been switched. Moreover, in each of the semesters students of female gender, non-Dutch ethnicity or high cognitive ability are not significantly more or less likely to be part of

[4] Lower dropout rates than the national average at the department of international business studies of the Amsterdam College of Applied Sciences can be explained by the fact that international programs generally attract more motivated students.

the management team than students of male gender, Dutch ethnicity or low cognitive ability. Thus, it is likely that the appointment of members to management teams also did not contaminate the initial gender, ethnic or ability composition of teams.

All in all, it is safe to assume that composition dynamics did not systematically alter the exogenously imposed team composition in the three field experiments.

2.3 Incentives

The entrepreneurship program provides various strong incentives to align the interests of students with the business performance of their company, both individually and at the team level.

As for individual incentives, students can be dismissed in case of underperformance. Dismissal of team members requires a two-third voting majority in the team together with the consent of the professor, and has severe consequences. Dismissed students are excluded from the program, lose its corresponding 12 credit points (out of 60 credit points in the first year) and endanger their prospect of obtaining an undergraduate degree (for which a minimum number of 45 credit points in the first year is mandatory). It is a credible threat. In the academic year of 2008-2009 half of the teams experienced at least one dismissal and the average number was 0.73 per team, while in the academic year of 2009-2010 nearly 30% of the teams experienced at least one dismissal and the average number of dismissals was 0.35 per team.

Another incentive with an individual component is the grade students obtain for the program from their professor. The program grade has a substantial weight of 20% in students' (first-year) grade point average. Both individual and team performance determine the program grade and their weights in the total program grade are about 50/50 (assessment of both components is based on the professor's subjective evaluation). Individual performance of students mainly entails active participation and the development of competencies such as cooperation, entrepreneurial behavior and professionalism. An indicator of the effect of individual performance on the program grade is the considerable average difference between the highest and lowest program grade within a team of about 1.5 in both academic years (on a scale from 1 to 10). The relevance of team performance for students' program grade is indicated by the positive correlations between teams' average program grade and their business outcomes.

The third individual incentive to align the interests of students with those of the team comes from the fact that virtually all students own one or more shares in their companies (with a nominal value of 20 euros per share). Roughly half of the shares are owned by team members themselves (approximately 50 euros per student on average); the remaining shareholders are usually family members, friends or acquaintances. The exact identity of shareholders per team is unknown.

Specific team incentives are provided by the formal business competition between teams. At the end of the program, six selected teams present their results in a 'business pitch' to a jury of entrepreneurs who choose a winner based on business performance and presentations. The winning team obtains a cup, often gets some (local) press attention and represents the college in the national Young Enterprise Start Up competition.

Reported effort levels in terms of hours are a quantitative reflection of the effectiveness of these incentives. On average, students spend about 8 hours per week on the program, which covers one-fifth of their first-year curriculum. This is substantial relative to the 32 hours per week students in Dutch higher vocational schools spend on their education (Allen et al., 2009). In sum, the incentives discussed above provide grounds to believe that the efforts of students are directed towards the business outcomes of teams.

Chapter 3

Gender diversity and team performance

This chapter is based on Hoogendoorn, S., H. Oosterbeek, and M. van Praag (2013), The impact of gender diversity on the performance of business teams: Evidence from a field experiment, Management Science, forthcoming.

3.1 Introduction

This chapter reports about a field experiment conducted to investigate the effect of gender diversity on team performance. The gender mix of a team may affect its performance through various underlying mechanisms. Hamilton et al. (2003) point to the trade-off between the costs of coordination and communication and the benefits from a potentially more diverse pool of knowledge and skills. The trade-off for these factors depends on the setting. In some contexts, coordination and communication costs are likely to be very high (the army). In other contexts, diversity of knowledge and skills may be very valuable (raising children).

Adams and Ferreira (2009) discuss 'mutual monitoring' as a mechanism and show that more gender diverse boards are associated with more intense monitoring practices (see also Gul et al., 2011). Dufwenberg and Muren (2006) derive results from a group dictator game played in the laboratory showing that mixed gender teams are more generous and more egalitarian. Woolley et al. (2010) show that teams with a larger percentage of women perform better due to a higher average level of social sensitivity of the group members. Pelled (1996) and Pelled et al. (1999) point to interpersonal aspects such as friendships and conflicts that vary with teams' gender composition and may affect their performance.

The results of our experiment among teams that have to start up and liquidate a real company may be informative about the effects of gender mix in teams that operate in comparable settings. One such setting, though admittedly not entirely comparable, are corporate boards. Recently the gender mix of these boards attracted considerable attention due to the current underrepresentation of women (see Catalyst, 2010; Woods, 2010) and the introduction of quota in some countries, including Norway, The Netherlands, Spain, France and Iceland. In 2010, women held only 10 percent of the board seats at the top 300 European companies and just above 15 percent of board seats at Fortune 500 companies (Catalyst, 2010; Woods, 2010). A higher share of women in boards is often regarded as desirable. Commonly expressed arguments in the popular press in favor of more gender diversity in corporate boards include: enlargement of the pool from which talent is attracted, complementarities and better mutual learning (Desvaux et al., 2007).[1] However, there is little empirical evidence supporting that gender diversity leads to better team performance.

Identifying the causal impact of gender diversity on the performance of teams is a challenging task. Studies based on observational data are likely to be plagued by various endogeneity issues, including reversed causality, unobserved heterogeneity and (self-)selection into and out of teams (Adams et al., 2010; Hermalin and Weisbach, 2003). We illustrate these issues by discussing some recent studies that acknowledge these measurement issues and address (part of) them.

Adams and Ferreira (2009) analyze data from US firms and find that the gender diversity of boards has a positive impact on the intensity of their mutual monitoring and the performance of firms that have otherwise weak (external) governance. However, more gender diverse boards are harmful for the performance of firms with strong (external) governance, possibly due to overmonitoring. On average, the effect of gender diversity on the firm performance indicators Tobin's q and ROA turns out to be negative. To address the possible endogeneity of gender diversity, they use the fraction of male board members with connections to female directors in other board positions as instrumental variable and include firm fixed effects in their regressions. Adams and Ferreira critically argue that these choices do not solve the issues of endogeneity and reverse causality completely: firm fixed effects only control for unobservables that are fixed over time and the instrument used is not so strong and may cause weak instrument problems. Moreover, it is questionable that the instrument passes the exclusion restriction required for an instrumental variable to be valid.

[1] In a laboratory experiment, Niederle and Vesterlund (2007) find that men often choose to compete even if they perform poorly while women often choose not to compete even if they perform well. If reaching a corporate board seat requires one to be competitive, these findings imply that the underrepresentation of women indeed reduces the pool from which talent can be attracted. Adams and Funk (2012) show that female board directors who break the glass ceiling are indeed a specific subset of females, in some respects more similar to males than the general population.

Ahern and Dittmar (2012), Matsa and Miller (2010) and Nygaard (2011) all measure the effect of board composition on firm performance and/or governance by exploiting that publicly listed firms in Norway were forced to have at least 40 percent female directors by 2008. In 2006, when this law was implemented, only 9 percent of directors were women. Firms thus had to replace on average 30 percent of their board members. In a difference-in-differences framework, Ahern and Dittmar compare before-after differences between early compliers and late compliers. Matsa and Miller compare listed and unlisted companies and companies in Norway and in other Scandinavian countries in a double and triple differences framework. Nygaard measures the stock market reaction to the unanticipated announcement of the quota as an indicator of the expected impact of an increase in the percentage of female directors on firm value and conditions on firm-specific information asymmetry. The three studies draw opposing conclusions. Ahern and Dittmar conclude that the reform had a significantly negative impact on firm value due to the fact that the newly added board members where younger and less experienced. Matsa and Miller are unable to distinguish between positive and negative effects on long term profits. Nygaard finds a significantly positive effect of the announcement of the law on the cumulative abnormal returns (CAR) for firms with low information asymmetry, while firms with high information asymmetry experience negative but insignificant CAR.[2]

Finally, Apesteguia et al. (2012) analyze data from the 2007-2009 editions of an online business game for students to study the effect of gender diversity on team performance. Almost 38,000 students in 16,000 teams participated. Incentives are strong: teams can win substantial prizes with relatively high probabilities, and there is the possibility of being hired by the company that runs this business game. The results show that teams of only women perform worse than mixed teams or teams of only men. However, there is no attempt (or mentioning) in this study to correct for the endogeneity of team formation. Teams that sign up for the game form themselves.[3]

Besides the studies mentioned, quite a number of field studies measure the relationship between teams' gender diversity and performance (e.g., Chowdhury, 2005; Ellison et al., 2010; Farrell and Hersch, 2005; Herring, 2009; Horwitz and Horwitz, 2007; Lee and Farh, 2004; Pelled, 1996; Pelled et al., 1999; Richard et al., 2004; Wegge et al., 2008). These contributions, however, study correlations rather than effects and do not

[2]Nygaard attributes the differences between his findings and those of Ahern and Dittmar to oversampling of new firms in the latter study.

[3]Related is also Hansen et al. (2006) who measure the impact of gender diversity in student groups on their grade for a group assignment that forms part of an undergraduate introductory management course. Male-dominated groups perform worse on a group-based performance measure than diverse groups and female-dominated groups. Performance in this study is academic achievement rather than business outcomes. Other studies looking at peer effects in education include Hoxby (2000), Lavy and Schlosser (2011) and Oosterbeek and Van Ewijk (2010).

aim at overcoming endogeneity or selectivity issues.[4] The effect of gender diversity in teams has also been studied in laboratory experiments (e.g., Dufwenberg and Muren, 2006; Ivanova-Stenzel and Kuebler, 2011; Pearsall et al., 2008). Experiments in the lab do not suffer from endogenous team composition but their resemblance to real-world situations may be limited, particularly for the topic at hand (see Chapter 1).

Our study is the first to conduct a field experiment with random assignment to circumvent endogeneity of teams' gender composition. The real life situation reflected in the experiment shares some features of corporate boards: the team size is comparable and so are the tasks. Moreover, female and male team members are of comparable quality in terms of education and experience (unlike in the Norwegian case).

Forty-three student companies are included in the experiment, with the majority of the observations with a share of women between 0.2 and 0.6. On this segment we find that teams with an equal gender mix perform better than male-dominated teams in terms of sales and profits. We do not have enough female-dominated teams to conclude that these results are symmetric. However, both the univariate and the regression analyses suggest that female-dominated teams do not perform better than gender diverse teams. We inquire various mechanisms suggested in the literature (including complementarities, learning, monitoring and conflicts) but fail to find support for any.

The remainder of this chapter is structured as follows. Section 3.2 describes the data and reports results from randomization checks. Section 3.3 presents the empirical findings. Section 3.4 discusses and concludes.

3.2 Data

In addition to administrative data and teams' annual reports, information was collected through three surveys. At the first day of the first week of the academic year (in September 2008), students filled out a pretreatment survey. Follow-up surveys were administered halfway (in January 2009) and at the end of the program (in May 2009). Response rates are 88% for the baseline survey, 86% for the first follow-up and 78% for the second follow-up. The surveys provide background information about individuals

[4] These studies that come from various literatures such as economics and finance (Ellison et al., 2010; Farrell and Hersch, 2005), organization and management (Horwitz and Horwitz, 2007; Pelled, 1996; Pelled et al., 1999; Richard et al., 2004), entrepreneurship (Amason et al., 2006; Chowdhury, 2005) or psychology (Lee and Farh, 2004; Wegge et al., 2008) consider various levels at which diversity and performance are measured, i.e., entire organizations, teams in the workplace, start-up or top management teams (boards). Numerous other field studies document the correlation between team diversity along various other dimensions than gender and team performance (e.g., Amason et al., 2006; Ancona and Caldwell, 1992; Bell, 2007; Edwards et al., 2006; Ensley and Hmieleski, 2005; Hamilton et al., 2012; LePine, 2003).

and teams. We use this information to assess whether team assignment was random conditional on gender (discussed in this section) and to inquire possible explanations for the effect of teams' gender mix on their business performance (see Section 3.3).

3.2.1 Gender diversity

Table 3.1 shows the numbers of students and teams, and share of women by field of study (teams in our experiment consist of students from the same field of study). The overall share of women is equal to 0.44 and the variation ranges from 0.1 to 1.0 in our sample. However, there are no teams with a share of women above 0.58 outside the field of business languages and there are no teams with a share of women below 0.17 outside the field of trade management Asia. This means that only the range from 0.17 to 0.58 is covered by more than one field of study. Of the 11 teams from the field of business languages, only 3 have a share of women below 0.58. Hence, the relation between share of women and performance is poorly identified in the range above 0.58. Our main focus in the rest of this chapter will therefore be on the majority of teams with a share of women between 0.2 and 0.6.

Table 3.1. Numbers of students and teams, and share of women by field of study

Field of study	Students	Teams	Share of women (per team)		
			Average	Minimum	Maximum
Business management	240	18	0.37	0.18	0.53
Management	60	5	0.29	0.17	0.40
Trade management Asia	105	9	0.35	0.10	0.58
Business languages	118	11	0.71	0.44	1.00
Financial management	27	2	0.40	0.38	0.42
Total	550	45	0.44	0.10	1.00

3.2.2 Other variables

The baseline survey contains questions about individual characteristics such as age, ethnicity, nationality, education and parental background. The average age is approximately 19 years and 4 months, roughly two-thirds of the population lives with their parents, about one-third has some work experience, and over 30% has a father who is or was an entrepreneur. Twenty percent of the students is born abroad and about half of the students has at least one parent not born in the Netherlands. The baseline survey also included the standard battery of questions to measure the five-factor model of personality structure known as the 'big five': agreeableness, conscientiousness, extroversion, neuroticism and openness to experience (see Goldberg, 1990). This commonly used set of measures of personality has been shown to be an explanatory factor of entrepreneurship choices and outcomes (Shane et al., 2010; Zhao and Seibert, 2006). Moreover, the baseline survey included statements that are combined through factor analysis into measures of entrepreneurial traits such as need for achievement, need for power, perseverance, risk aversion, self-efficacy and social orientation. These traits are supposed to be constant over time and possibly influential for entrepreneurship decisions and outcomes (see Oosterbeek et al., 2010; Parker, 2009).

Table 3.2. Descriptive statistics of team characteristics

	Scale	Mean	SD	Min	Max
Team characteristics					
Age	years	19.37	0.58	18.29	20.93
Ethnicity	0-1	0.55	0.16	0.20	0.90
Nationality	0-1	0.21	0.11	0.00	0.50
Grade point average	1-10	6.46	0.07	6.33	6.64
Size (at baseline)	persons	12.22	2.09	8.00	16.00
Conflicts	1-5	2.23	0.59	1.00	3.67
Atmosphere	1-5	3.53	0.55	2.33	4.83
Peer-reviewed efforts	1-10	6.94	0.56	6.14	9.17
Dismissals (dummy = 1 if any)	0/1	0.49	0.50	0.00	1.00
Big five characteristics	Cronbach's α				
Agreeableness	0.75	6.39	0.60	5.13	8.03
Conscientiousness	0.77	3.08	0.66	1.58	5.00
Extroversion	0.81	-1.81	0.67	-2.85	0.99
Neuroticism	0.76	5.48	0.57	4.28	6.50
Openness to experience	0.63	7.35	0.46	6.30	8.21
Team processes					
Group potency	0.87	10.85	1.51	8.29	15.17
Decision making	0.70	1.58	1.24	-0.73	4.60
Mutual monitoring	0.88	9.99	1.02	7.83	12.69
Coordination	0.80	1.83	1.11	-0.31	4.90
Credibility	0.66	2.87	0.61	1.41	4.84
Specialization	0.66	7.88	0.60	5.97	9.66

Unlike these traits, entrepreneurial skills can be developed over time. Therefore, validated batteries of questions to measure the most relevant skills for entrepreneurship are included in all three surveys. The skills that are measured include analyzing skills, creativity, external orientation, flexibility, market awareness, motivating skills, networking skills, organizing skills and pro-activity (see Parker, 2009). These measures are taken using the so-called Escan, a validated self-assessment test based on 114 items. The Escan is widely used in the Netherlands to determine people's entrepreneurial competencies by, for instance, the Dutch Chambers of Commerce and commercial banks. The statements load into factors (with Cronbach's alphas ranging from 0.64 to 0.79) of which the entrepreneurship literature has shown to be the most important traits and skills for successful entrepreneurship. Based on the data collected in Oosterbeek et al. (2010) it has been slightly adapted to increase the validity of items when a population of students rather than entrepreneurs is involved. Finally, all three surveys include self-assessments of the knowledge that students have in seven areas that are relevant for entrepreneurship, i.e., knowledge of business, management, entrepreneurship, strategy, organization, administration and leadership (see Karlan and Valdivia, 2011; Minniti and Bygrave, 2001).

To help explain possible differences in team performance, the second follow-up survey contains questions related to teams' procedures and processes. We obtain measures of the teams' atmosphere, conflicts, peer-reviewed individual effort, friendships, dismissals, satisfaction and the existence of subgroups. Questions related to processes within the team translate into measures of group potency (De Jong et al., 2005), decision making (Oliver and Anderson, 1994), mutual monitoring (Langfred, 2004) and coordination, credibility and specialization (Lewis, 2003). Table 3.2 reports the scales on which these variables are measured and descriptive statistics at the team level. This table shows that there is quite some variation in the scores on these variables across teams. In Section 3.3 we examine to what extent these scores are related to teams' gender composition to verify whether they can potentially explain our results.

The outcome variables in our analyses are measures of teams' sales and profits. Information about these variables was retrieved from the annual reports that we managed to obtain from 43 out of 45 teams. Sales measures are straightforward and uniformly reported in these reports. Obtaining comparable profit numbers is more challenging and required careful examination of the students' financial statements. For instance, the way the wages of team members, depreciation and the costs of unsold goods were accounted for was not always uniform. We corrected this as much as possible and are confident that the profit measure is more noisy than the sales measure, but not more noisy than profit measures in other datasets.[5]

The first column of Table 3.3 shows descriptives of sales and profits based on the information from annual reports of all 43 teams. Average sales add up to 838 euros, with a standard deviation of 707 euros. The worst performing team has no sales, while the best performing team sells for more than 4000 euros. Profits are on average negative at -69 euros. The team with the lowest profits loses 1016 euros, while the highest profits are 477 euros. The correlation between sales and profits equals 0.25 (p=0.11). The next three columns break the descriptives of the performance measures down by three groups of teams; teams with a low share of women (less than 0.4), teams with a moderate share of women (between 0.4 and 0.6) and teams with a high share

[5]Unlike in other studies using samples of firms and their profit numbers, all companies produced their financial report in the same format and we were allowed to check each entry.

of women (above 0.6). The consistent picture emerging from this breakdown is that on average sales and profits are higher for teams with a balanced gender mix than for teams dominated by one sex. This alludes to the main finding of this chapter, which we will present more formally in Section 3.3.

Table 3.3. Descriptive statistics of outcome variables, in euros

			Share of women (per team)		
		All	Low	Moderate	High
			I[share<0.4]	I[0.4≥share≤0.6]	I[share>0.6]
Sales	Mean	838	698	1209	609
	SD	707	441	1023	506
	Min	0	0	265	205
	Max	4209	1953	4209	1618
Profits	Mean	-69	-104	17	-116
	SD	318	325	249	413
	Min	-1016	-1011	-402	-1016
	Max	477	477	294	152
N		43	23	13	7

3.2.3 Randomization

Before we get to the main results, we first examine whether students are randomly assigned to teams of different gender composition, conditional on their gender. We test this by regressing - separately for male and female students - students' characteristics on the share of women in their team. Since in the next section we allow for non-monotonic patterns in the relation between team performance and the share of women, we also do that in the randomization checks. We regress the characteristics of students on dummies for the share of women in the team being below 0.4 or above 0.6 (results are very similar when we employ other cutoffs such as 0.33 and 0.67 or 0.45 and 0.55). The reference category is a balanced gender mix between 0.4 and 0.6.

Table 3.4 reports the results. Each pair of coefficients comes from a separate regression. For example, we regressed age of male students on dummies for share of women below 0.4 and above 0.6, and find coefficients equal to 0.166 (s.e. 0.330) and 0.783 (s.e. 0.497). Ideally, none of the coefficients of the table should be significantly different from zero. There are some deviations from this ideal. Men assigned to groups with a low share of women are less open to new experience than men assigned to a team with a balanced gender mix. Likewise, men assigned to groups with a high share of women are more likely to be of Dutch origin and score lower on conscientiousness than men assigned to a team with a balanced gender mix. Finally, women assigned to groups with a high share of women are a bit older and score higher on conscientiousness than women assigned to a team with a balanced gender mix. Note that four of the five significant coefficients pertain to teams with a high share of women. As we mentioned above, this is the segment where the effect of gender composition on team performance is poorly identified, implying that we cannot draw firm conclusions about these effects within this segment anyway.

Table 3.4. Random assignment of male and female students at the individual level

	Males		Females	
	I[share<0.4]	I[share>0.6]	I[share<0.4]	I[share>0.6]
Personal characteristics				
Age	0.166	0.783	0.315	0.652**
	(0.330)	(0.497)	(0.348)	(0.315)
Ethnicity	-0.070	-0.330**	0.045	-0.112
	(0.055)	(0.142)	(0.094)	(0.113)
Nationality	-0.026	-0.025	0.016	-0.045
	(0.060)	(0.222)	(0.062)	(0.080)
Grade point average	0.009	0.054	-0.017	0.050
	(0.026)	(0.061)	(0.056)	(0.065)
Big five characteristics				
Agreeableness	0.066	-0.671	0.008	-0.084
	(0.265)	(0.558)	(0.327)	(0.379)
Conscientiousness	0.001	-1.169**	0.224	0.665***
	(0.272)	(0.537)	(0.290)	(0.165)
Extroversion	0.423	0.139	-0.176	-0.308
	(0.335)	(0.432)	(0.270)	(0.341)
Neuroticism	-0.404	-0.034	-0.100	0.669
	(0.274)	(0.589)	(0.225)	(0.449)
Openness to experience	-0.400**	-0.008	0.197	0.112
	(0.169)	(0.392)	(0.178)	(0.381)

Note: Pairs of coefficients come from a regression at the individual level of the row variable on dummies for share of women below 0.4 and above 0.6, separately for men and women. All regressions include controls for field of study. Robust standard errors in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

The share of women is possibly correlated with other team characteristics. Table 3.5 reports results from regressions of team characteristics on dummies for low and high share of women. Team characteristics are the mean values of the individual characteristics analyzed in Table 3.4. None of the team characteristics is significantly related to the share of women. The coefficients of a low share of women are relatively precisely estimated. The coefficients of a high share of women are not precisely measured. This is due to the small number of observations in that category.

Table 3.5. Regressions of (average) team characteristics at baseline on dummies for share of women

	I[share<0.4]		I[share>0.6]	
Personal characteristics (average)				
Age	0.233	(0.215)	0.707	(4.060)
Ethnicity	-0.030	(0.068)	-0.201	(1.006)
Nationality	-0.020	(0.039)	-0.111	(0.451)
Grade point average	0.006	(0.031)	0.048	(0.342)
Big five characteristics (average)				
Agreeableness	-0.041	(0.199)	-0.472	(1.336)
Conscientiousness	0.010	(0.238)	0.015	(2.442)
Extroversion	0.280	(0.212)	-0.055	(1.183)
Neuroticism	-0.191	(0.189)	0.356	(3.668)
Openness to experience	-0.026	(0.135)	-0.321	(2.549)
Team size	0.412	(0.721)	0.021	(8.335)

Note: Pairs of coefficients come from a regression at the team level of the row variable on the column variables. All specifications include controls for field of study. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

3.3 Results

3.3.1 Main findings

Figures 3.1 and 3.2 show the relations between the share of women in a team and teams' sales and profits. The graphs are based on kernel-weighted local polynomial smoothing (details are reported below each graph). Dots represent the actual team results and the shaded areas the 90% confidence intervals. The relation between sales and share of women is inverse u-shaped. For a share of women between 0.2 and 0.5, sales increase when the share of women increases. When the share of women exceeds 0.5, sales tend to decrease when the share of women increases further. Also profits are increasing in the share of women when the share of women is below 0.5. For higher shares of women, the relation between profits and the share of women is flat. The dots show clearly that almost all best-performing teams have an equal gender mix, while teams that perform poorly are more spread out across the distribution.

Tables 3.6 and 3.7 report results from different regressions of sales and profits on the share of women. Motivated by the graphs we divide the horizontal axis into three segments: a segment with a low share of women, a segment with a balanced gender mix and a segment with a high share of women. We present results for three pairs of cutoffs

Figure 3.1. Relation between share of women and sales

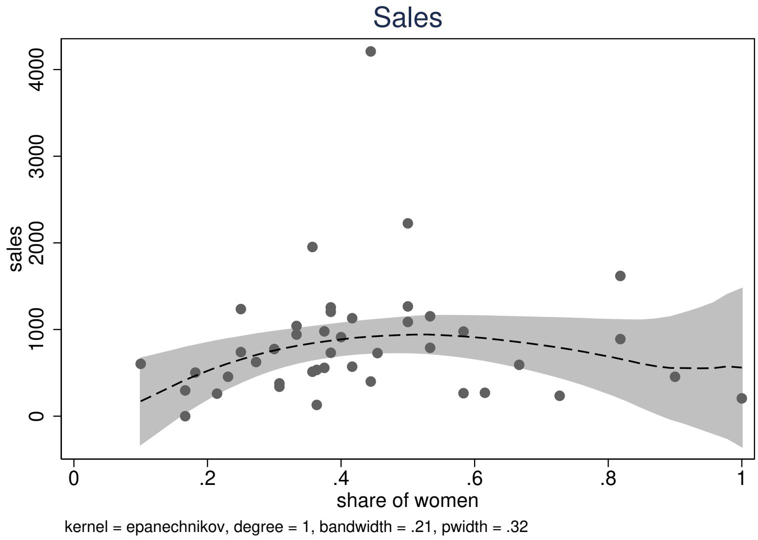


Figure 3.2. Relation between share of women and profits

between low-balanced and balanced-high: 0.4 and 0.6; 0.45 and 0.55; 0.49 and 0.51.[6] In panel A, the relation between outcomes and share of women is captured in splines. This allows for different linear relations between the outcomes and the share of women on each segment. In panel B, we collapse the share of women into two dummy variables for a low and a high share of women (where a balanced share is the reference group). Columns (1), (4) and (7) report results from standard OLS regression. Since these results may be sensitive to outliers we also present results from median regression in columns (2), (5) and (8), and from robust regression (using M-estimation) in columns (3), (6) and (9). Since the number of observations is small, analytical standard errors based on asymptotic theory may understate and we therefore report bootstrapped standard errors based on 1000 replications.[7]

Results in panel A of Table 3.6 show that sales are increasing in the share of women in the first segment. For the second and third segments none of the estimates is significantly different from zero. Results are not very sensitive to the precise location of the first spline point. Comparing the OLS results for the first segment with the results from median and robust regression shows that the OLS estimates are substantially inflated by outliers; the estimates are more or less cut in half when we move from OLS to the other estimation methods. The result in column (2) of panel A implies that raising the share of women from 0.3 to 0.4 increases sales by 225 euros (about one-third of a standard deviation).

Panel A of Table 3.7 indicates that profits are also increasing in the share of women in the first segment. However, not all estimates are precise enough to reach statistical significance. The point estimates are very stable across the location of the spline point and the estimation method. Also for profits, none of the splines on the second and third segments are statistically significant.

The results from the dummy regressions in panels B are in accordance with the results in panels A. Teams with a low share of women have lower sales and lower profits than teams with a balanced gender mix. We also find that teams with a large share of women make lower profits than teams with a balanced gender mix. This finding is, however, sensitive to the definition of the dummy variables. It matters a lot for the estimates whether some teams with a share of women in the range between 0.55 and 0.6 are assigned to the high or balanced group.

All in all, the results presented in this subsection indicate that teams' sales and profits increase when the share of women increases from a low to a moderate level.

[6]Measures of goodness of fit of these different models are very similar. The last pair of cutoffs leaves only three out of 43 teams in the balanced category. Results are very similar when we choose only one breakpoint at 0.5.

[7]We note that analytical standard errors are indeed substantially smaller than the bootstrapped standard errors reported here.

Table 3.6. Effect of share of women on sales; various specifications

	0.4 and 0.6			0.45 and 0.55			0.49 and 0.51		
	OLS	Median	Robust	OLS	Median	Robust	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A: Splines									
1st segment	3243.3**	2255.0*	2116.2**	3869.3**	2158.5*	1908.1**	3831.4**	2255.0*	1868.0**
	(1310.4)	(1343.8)	(1065.4)	(1829.6)	(1191.9)	(851.5)	(1833.1)	(1161.8)	(911.9)
2nd segment	1203.4	-356.5	570.7	-2801.0	-1669.7	531.5	-22604.2	-13168.6	-846.8
	(2193.6)	(2682.5)	(5052.0)	(4864.0)	(4270.8)	(4205.6)	(26382.9)	(20794.8)	(21157.6)
3rd segment	431.4	367.4	87.7	834.3	599.2	15.5	788.9	552.3	2.1
	(1691.5)	(2969.4)	(2640.4)	(1592.8)	(2027.5)	(2832.3)	(1682.5)	(2011.0)	(4652.8)
R^2	*0.22*	*0.21*	*0.33*	*0.25*	*0.22*	*0.35*	*0.25*	*0.22*	*0.35*
B: Dummies									
1st segment	-648.2*	-287.0	-309.1	-323.7	-491.7	-382.4	-720.8**	-507.9	-872.6**
	(347.3)	(245.4)	(218.4)	(316.6)	(310.2)	(260.3)	(357.8)	(478.7)	(428.6)
3rd segment	276.9	189.7	98.8	-165.9	-290.2	-219.8	-602.8*	-290.2	-603.5
	(208.7)	(235.0)	(211.2)	(315.4)	(342.9)	(289.8)	(363.6)	(494.5)	(437.6)
R^2	*0.23*	*0.19*	*0.27*	*0.10*	*0.17*	*0.26*	*0.14*	*0.18*	*0.34*

Note: Based on information from 43 teams. All specifications include controls for field of study. Numbers in the first row refer to cutoffs of the share of women used to create splines and dummies. OLS, Median and Robust refer to the estimation method. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

Table 3.7. Effect of share of women on profits; various specifications

	0.4 and 0.6			0.45 and 0.55			0.49 and 0.51		
	OLS	Median	Robust	OLS	Median	Robust	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A: Splines									
1st segment	750.7	954.2	682.8	840.7	814.2*	720.3	845.2	977.8**	731.2*
	(784.0)	(654.0)	(787.1)	(644.3)	(447.6)	(522.3)	(610.8)	(424.4)	(440.9)
2nd segment	-292.5	-934.2	-227.7	-1367.6	-2495.9	-1102.9	-8429.3	-18845.2	-7245.9
	(1023.8)	(1313.4)	(1135.7)	(1815.3)	(2262.6)	(2401.7)	(9222.2)	(11476.9)	(11019.4)
3rd segment	373.4	603.9	221.2	368.3	863.2	264.0	307.6	1034.4	224.4
	(1132.0)	(1563.1)	(2596.1)	(978.2)	(1318.0)	(2300.8)	(1070.2)	(1360.8)	(2141.2)
R^2	*0.22*	*0.17*	*0.25*	*0.23*	*0.19*	*0.26*	*0.23*	*0.20*	*0.26*
B: Dummies									
1st segment	-111.9	-80.0	-61.2	-160.7	-223.7*	-158.3	-214.0**	-223.7*	-188.3**
	(92.4)	(123.3)	(105.4)	(110.6)	(122.6)	(106.4)	(88.8)	(116.2)	(84.1)
3rd segment	-24.5	286.5	127.3	-301.3**	-365.1**	-237.3**	-268.2*	-365.1**	-253.7*
	(201.7)	(201.5)	(186.7)	(140.8)	(144.1)	(108.0)	(161.3)	(183.2)	(150.9)
R^2	*0.21*	*0.14*	*0.26*	*0.23*	*0.18*	*0.31*	*0.22*	*0.17*	*0.27*

Note: Based on information from 43 teams. All specifications include controls for field of study. Numbers in the first row refer to cutoffs of the share of women used to create splines and dummies. OLS, Median and Robust refer to the estimation method. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

Our estimates lack the precision to draw firm conclusions about the effect on business performance of the share of women in a team when this share exceeds 0.6. However, the results suggest that female-dominated teams do not perform better than gender diverse teams.

3.3.2 Mechanisms

The studies mentioned in the introduction of this chapter suggest various mechanisms as possible explanations for our findings. In what follows, we explore the potential of these mechanisms by regressing indicators of these mechanisms on the share of women in the team, thereby using the same specifications as before (with cutoffs at 0.4 and 0.6). If the variable that indicates a certain mechanism is unrelated to the share of women then we can safely conclude that this mechanism cannot explain our findings. If the variable that indicates a certain mechanism is related to the share of women then this mechanism potentially contributes to the explanation of our results. For the mechanism to actually explain our findings (partially), it also needs to be true that the variable that captures the mechanism has a significant impact on sales and/or profits. Unfortunately, our research design (nor any other we know of) does not allow us to test this. The variable of interest is endogenous and we lack exogenous variation, other than the randomization of the share of women, to identify its causal impact. Including the regressor of interest as an additional control next to the share of women introduces a 'bad controls' problem and renders the coefficients uninterpretable (see Angrist and Pischke, 2008). Results are reported in Table 3.8. The remainder of this subsection describes the operationalization of the mechanisms and discusses the results.

Complementarities

Men and women in mixed teams may complement each others' skills and knowledge. We standardize the various skill and knowledge dimensions (see Section 3.2) and then compute for each skill and knowledge dimension the maximum in a team. Subsequently, we compute the minimum of the maximums of all skill dimensions and the same for all knowledge dimensions. Supposedly, if men and women complement each others skills or knowledge, these minimums are higher in mixed teams. We find no support for that; see columns (1) to (6) of Table 3.8.

Learning

When teams learn, mean skill and knowledge levels increase. Learning may be related to the gender composition of a team. This may be due to different initial distributions of skills and knowledge levels or differential team processes that may be unobserved.

Table 3.8. Mechanisms

	Compl. knowledge			Compl. skills			Learn. knowledge			Learn. skills			Team conflicts			Friendships		
	OLS	Med	Rob	OLS	Med	Rob	OLS	Med	Rob	OLS	Med	Rob	OLS	Med	Rob	OLS	Med	Rob
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
A: Splines																		
1st segment	0.2	0.0	0.2	0.4	-0.3	0.4	-0.8	-1.1	-1.2	-0.0	-2.1	-0.1	-0.5	0.1	-0.3	-0.5	0.7	-0.2
	(1.0)	(1.7)	(1.2)	(1.1)	(1.9)	(1.4)	(0.9)	(1.2)	(0.9)	(1.1)	(2.1)	(1.7)	(1.4)	(2.1)	(2.1)	(1.4)	(1.9)	(1.7)
2nd segment	1.3	-0.0	1.3	-0.4	-0.3	-0.3	0.7	1.7	1.8	-0.3	0.7	-0.0	1.0	-0.4	0.8	4.6*	3.3	4.5
	(2.0)	(3.0)	(2.6)	(1.8)	(2.9)	(2.6)	(1.4)	(1.6)	(1.1)	(1.9)	(3.1)	(2.8)	(2.5)	(4.3)	(4.1)	(2.4)	(3.3)	(3.0)
3rd segment	0.7	0.0	0.8	-1.9	-0.7	-1.7	-1.7**	-1.6	-1.9	0.7	0.7	0.7	-0.3	-1.0	-0.4	-2.5	-4.0	-2.6
	(1.8)	(2.3)	(6.5)	(3.2)	(2.5)	(6.5)	(0.8)	(1.5)	(1.4)	(1.4)	(3.6)	(3.3)	(4.0)	(2.0)	(6.5)	(1.7)	(2.8)	(6.8)
B: Dummies																		
1st segment	-0.1	-0.0	-0.0	-0.0	0.1	-0.0	0.1	0.0	0.0	0.1	0.1	0.1	-0.1	-0.0	-0.1	-0.2	-0.3	-0.3
	(0.2)	(0.3)	(0.2)	(0.1)	(0.2)	(0.2)	(0.1)	(0.1)	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)	(0.3)	(0.3)	(0.2)	(0.2)	(0.2)
3rd segment	0.4	0.0	0.4	0.0	-0.1	0.1	-0.2	-0.6*	-0.2	0.1	-0.0	0.0	-0.1	0.2	-0.1	0.1	-0.9	0.1
	(0.3)	(0.4)	(0.3)	(0.3)	(0.4)	(0.4)	(0.2)	(0.3)	(0.3)	(0.2)	(0.3)	(0.2)	(0.2)	(0.3)	(0.3)	(0.7)	(1.0)	(0.8)

	Decision making			Team atmosphere			Low / high tech	Product / service	Monitoring		
	OLS	Med	Rob	OLS	Med	Rob	OLS	OLS	OLS	Med	Rob
	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)
A: Splines											
1st segment	2.1	1.4	1.2	1.6	1.0	1.3	1.1	0.3	2.4	2.6	3.2
	(2.8)	(4.2)	(3.6)	(1.3)	(1.7)	(1.7)	(1.2)	(0.9)	(2.5)	(3.4)	(3.2)
2nd segment	-3.0	-1.7	-1.6	-0.1	2.0	0.9	-2.0	1.0	3.3	4.8	2.7
	(5.9)	(8.9)	(7.9)	(2.3)	(3.6)	(3.3)	(2.3)	(1.1)	(3.6)	(4.8)	(4.5)
3rd segment	2.9	2.8	3.3	1.2	1.1	1.3	-0.3	1.1	0.2	3.2	0.9
	(4.5)	(5.9)	(8.4)	(1.6)	(1.8)	(3.2)	(1.5)	(1.2)	(3.3)	(4.3)	(5.6)
B: Dummies											
1st segment	0.4	-0.1	0.3	-0.0	-0.0	-0.1	0.1	-0.0	-0.4	-0.5	-0.4
	(0.5)	(0.7)	(0.6)	(0.2)	(0.3)	(0.2)	(0.2)	(0.1)	(0.3)	(0.5)	(0.5)
3rd segment	0.4	1.1	0.6	0.2	0.3	0.3	0.1	-0.1	0.8	0.5	0.9
	(0.7)	(0.7)	(0.7)	(0.2)	(0.3)	(0.3)	(0.2)	(0.2)	(0.6)	(0.7)	(0.7)

Note: Based on information from 43 teams. All specifications include controls for field of study. OLS, Median and Robust refer to the estimation method. Median and robust specifications for type of product are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

The team average increases in skill/knowledge levels turn out to be unrelated to teams' gender composition; see columns (7) to (12) of Table 3.8. Thus, there is not more or less (mutual) learning in gender diverse teams than in other teams.

Conflicts, friendships, decision making and atmosphere

The second follow-up survey asked to what extent there was conflict or disagreement between the team members about personal matters (that did not have anything to do with performing the tasks). Examples are social events or gossip. Respondents could give a score on a scale from 1 to 5. The average score of a team on this variable is unrelated to teams' gender composition; see columns (13) to (15). The same holds for agreeableness or social skills.

Moreover, in the second follow-up survey, we asked respondents whether decisions on strategies were mainly taken by a few members of the team or were generally taken by the whole team. Teams' averages of this variable are unrelated to gender diversity; see columns (19) to (21). The second follow-up survey also asked respondents how many team members they see on a friendly basis. We took the average of that number as an indicator of friendships in a team. This measure is unrelated to gender diversity; see columns (16) to (18). Finally, the second follow-up survey asked respondents to rate the atmosphere within their team on a 5-points scale. The average within a team is our measure of team atmosphere. The atmosphere within teams turns out to be unrelated to gender diversity; see columns (22) to (24).

Type of product

We have also checked whether the products/services produced by more gender diverse teams target a more diverse and thus larger market. To this end, we have categorized the companies and their products in various ways. The market orientation of teams is as follows: 40% focuses on the looks of customers, 33% employs high technology, 88% sells a product (rather than a service), and 23% exclusively targets female buyers. No systematic relationships were observed between the market orientation of teams and their gender mix; see columns (25) and (26).

Monitoring

We collected information to measure the level of mutual monitoring in both follow-up surveys. The measure of monitoring is based on four items (see Langfred, 2004): (i) we check to make sure that everyone in the team continues to work; (ii) we check whether everybody is meeting their obligations to the team; (iii) we monitor each other's progress on the project; (iv) we watch to make sure that everyone in the team meets their deadlines. The Cronbach's alpha of 0.88 indicates the validity of the factor. We restrict the analyses to the level of monitoring measured in the second follow-up (in May 2009) because students are likely to have a more comprehensive overview of mutual monitoring in their teams at the end of the program. We find no evidence that monitoring varies with the share of women in the team; see columns (27) to (29).

3.4 Discussion and conclusion

The key finding of this study is that business teams with an equal gender mix perform better than male-dominated teams in terms of sales and profits. Our study also suggests that teams with an equal gender mix perform no worse than teams with a majority of females, although the distribution of our data does not allow firm conclusions about the effect of gender mix on performance for female-dominated teams.

Some important things remain unknown. Practical matters restricted the range over which we could vary the share of women. Consequently, the results are most informative about the effects of changing from a male-dominated team into a team with a balanced gender mix. The results are less informative about the effects of changing from a female-dominated team into a team with a balanced gender mix. Nevertheless, the range of the gender distribution where we have sufficient variation is relevant for feeding the current discussion of 'improving' the gender balance in management teams. While we collected detailed information about team processes, we were unable to identify any process that could potentially explain why mixed teams perform better than male-dominated teams.

The entrepreneurship program in which we were allowed to manipulate the gender composition of teams is used in many schools in many countries. This gives ample opportunity for follow-up studies. These studies should in the first place attempt to replicate results from the current study, preferably on a larger scale. If possible, these studies should also cover a wider range of the share of women, so that we can learn whether effects are symmetric around a balanced gender mix. Finally, with research on a larger scale, it should also be possible to uncover some of the underlying mechanisms.

Chapter 4

Ethnic diversity and team performance

This chapter is based on Hoogendoorn, S. and M. van Praag (2012), Ethnic diversity and team performance: A field experiment, Tinbergen Institute Discussion Paper No. 2012-068/3.

4.1 Introduction

We measure the causal effect of ethnic diversity on the performance of business teams using a field experiment. The choice for this topic and approach are easily motivated. One of the most salient dimensions of team heterogeneity is ethnicity (Alesina and La Ferrara, 2005). Ethnic diversity implies heterogeneity in (mother) languages, religions, races and cultures (Alesina and La Ferrara, 2005). It is commonly measured based on country of birth, of the individual or of his/her parents. Ethnic diversity also coincides with a variety of norms, information sets, knowledge and ability levels (Lazear, 1999; Morgan and Vardy, 2009). This variety affects the formation and performance of teams.

Ethnic diversity would benefit team performance due to a more diverse pool of skills and knowledge that leads to complementaries and (mutual) learning. For example, due to complementarities and learning opportunities, ethnically diverse teams are associated with more creativity and innovation (Alesina and La Ferrara, 2005; Lee and Nathan, 2011; Ozgen et al., 2011b). On the other hand, the costs associated with more ethnic diversity would be related to more difficult communication and coordination (Lazear, 1999; Morgan and Vardy, 2009).[1] All in all, ethnic diversity is an influential source of heterogeneity.

[1]Ethnic diversity may also affect group formation and performance through its influence on the

Ethnic diversity is highly relevant in an increasingly globalized world. Multinational firms often staff teams internationally and local populations - especially in big cities - become more mixed and multicultural. Ethnic diversity is a current fact of life and the share of ethnic minorities in Western populations is increasing sharply (Alesina and La Ferrara, 2005; Lazear, 1999; Ozgen et al., 2011b). In the United States, for example, the share of minorities is expected to rise from about one-third nowadays to roughly the majority in 2042 (Bernstein and Edwards, 2008). As a consequence, it is likely that any team will become more and more diverse in terms of ethnicity, even if the optimal team formation would indicate otherwise.

The relevance and potential impact of ethnic diversity in teams motivate our choice for the topic of this study. Our approach of a field experiment among business teams that start up in identical circumstances is motivated based on the literature. Team formation is obviously driven by prospective productive consequences. If the situation carries a higher likelihood that an ethnically diverse team is beneficial, the team composition will be more mixed (Boisjoly et al., 2006). Hence, the measured effects of ethnic diversity on performance in real-world teams are likely to be biased due to endogenous team composition.

Examples of studies conducted on the ethnic diversity of teams in real organizations include Carter et al. (2010), Hamilton et al. (2012), Kahane et al. (2013), Leonard et al. (2010) and Parrotta et al. (2010), and their results are ambiguous (Alesina and La Ferrara, 2005).[2] Experiments in the laboratory have established results that are largely consistent with the theory proposed by Lazear (1999). An optimal degree of heterogeneity results from the trade-off between the benefits of more ethnic diversity and the associated increased costs of communication and coordination (Alesina and La Ferrara, 2005).

Some studies measuring the effect of ethnic diversity have tried to combine the advantages of studies in real organizations with experimental studies by carrying out field or 'quasi' experiments. Hansen et al. (2006) resemble the design of a field experiment. They measure the impact of demographic diversity (age, gender and ethnicity) in student groups of four to five students on the team's academic performance and find no effect.[3] Boisjoly et al. (2006) find that attitudes and behaviors change when people

group culture (Earley and Mosakowski, 2000; Richard et al., 2004) and the strategic behavior of group members (Alesina and La Ferrara, 2005).

[2]Other studies examine the effects of ethnic diversity on productivity at the country (Guiso et al., 2009; Montalvo and Reynal-Querol, 2005), region (Ozgen et al., 2011a) or city level (Lee and Nathan, 2011; Ottaviano and Peri, 2006). Related are also studies measuring the impact of ethnic composition of schools or neighborhoods on educational outcomes (e.g., Angrist and Lang, 2004; Aslund et al., 2011; Card and Rothstein, 2007; Hanushek et al., 2009; Hoxby, 2000).

[3]Group composition is random and no exogenous stratification is imposed. Teams are required to select one of three contract forms that determine the authority of grading. The drawback of this design, when interested in the bare effect of ethnic diversity on team performance, is that the effect

of different ethnicity are randomly assigned to live together at the start of their first year of college. White students assigned to African-American roommates show to be significantly more empathetic to these groups.

We conclude that measurements of the causal effect of a team's ethnic diversity on its performance are scarce, especially in the longer run. This kind of measurement is the objective of our study. We exploit the fact that the - otherwise homogeneous - population consists of 55% students with a non-Dutch ethnicity from 53 different countries of origin and that we, as outside researchers, were allowed to exogenously vary the ethnic composition of otherwise randomly composed teams. The resulting percentage of students with a non-Dutch ethnicity per team varies from 20% to 90%. We use a rather common definition of ethnicity, i.e., parents' country of birth.

What kind of results might we expect? We combine Lazear's argument (1999) that there is a trade-off between the costs and benefits of an ethnically more diverse team with recent insights from Earley and Mosakowski (2000) and Richard et al. (2004) based on Blau's theory of heterogeneity (1977). The non-formal models in these studies refine Lazear's argument by allowing the costs of communication and coordination to be a non-linear function of ethnic diversity. Moderately heterogeneous teams would incur higher costs of communication and coordination than teams that are homogeneous or heterogeneous in terms of ethnicity. In these moderately heterogeneous teams subgrouping along ethnic lines may have negative effects such as distrust, conflicts or stereotyping between distinct subgroups.[4]

The non-linear relationship between communication costs and ethnic diversity could perhaps be illustrated by considering three types of team compositions: (i) with a low percentage of minorities, (ii) with a substantial subgroup of minorities, and (iii) with a majority group of minorities. In the first situation, the communication costs are probably low. The few minorities present will perhaps not participate in the team process and be left aside. This is costly in the sense that their productivity is lost, but communication can be low cost too and based on habits, language and norms of the majority group (i.e., with a single identity).

In the second situation, the group that may not take part in the productive process of the majority is larger. Segregated subgroups may be formed by the majority and the others with distinct manners of work. Hence, costs of diversity are higher at this stage. In the third situation, in a truly ethnically diverse team, the costs of communication

of interest may be confounded by the contract choice (that may be related to ethnic diversity).

[4]Akerlof and Kranton (2005) show that teams' identity could act as a substitute mechanism for formal incentive schemes. In a series of laboratory experiments, Charness et al. (2007) find that group membership and identification with this group, influence individual behavior in strategic environments. Moreover, individuals tend to operate significantly more altruistically towards ingroup members than to outgroup individuals (Chen and Li, 2009). In a field experimental setting, Dugar and Shahriar (2012) find that group identity can reduce moral hazard problems.

have probably become lower due to the absence of subgroups and the desirability for every team member to participate in a hybrid team culture with a diverse identity (Earley and Mosakowski, 2000; Richard et al., 2004).

We thus expect that the costs of ethnic diversity follow an inverse u-shaped pattern, whereas the benefits of complementarities and (mutual) learning may be an(y) increasing function of ethnic diversity (Lazear, 1999). However, due to the fact that we do not have any conjectures about the specific forms of the cost and benefit functions relative to each other, the net effect of ethnic diversity on team performance remains a question that needs to be answered empirically.

Our empirical analysis renders the following results. The impact of a team's ethnic diversity on its performance as measured in terms of business outcomes is positive, although only starting at a certain turning point at which at least half of the team is ethnically diverse. Before this turning point the relationship is flat or slightly negative. With respect to underlying mechanisms, our data show that more (mutual) learning could explain why ethnic heterogeneous teams achieve better results. Ethnically diverse teams also tend to have a more diverse pool of relevant knowledge facilitating (mutual) learning.

In what follows, Section 4.2 describes the data and presents results from randomization checks. In Section 4.3 we show the empirical findings. Section 4.4 provides a discussion and conclusion.

4.2 Data

Data about teams and individuals was collected through administrative data, teams' annual reports and three extensive surveys. Students filled out a survey at the first day of the academic year (pretreatment, in September 2008), halfway (in January 2009) and at the end of the entrepreneurship program (posttreatment, in May 2009). Response rates are respectively 88%, 86% and 78%. We use the information from the questionnaires for various purposes.

Based on this, we check the validity of the name-based assignment of students to ethnicities and the random assignment of students to teams given their ethnic background (see Chapter 2 and this section). Moreover, the questionnaires enable us to relate our findings to and control for individual and team characteristics and processes regarding complementarities, (mutual) learning, communication and coordination. As we will discuss more thoroughly in Section 4.3, communication and coordination are proxied by self-assessments of team members about the team's atmosphere and the prevalence of conflicts.

4.2.1 Ethnic diversity

Students are defined as native or Dutch if both parents are born in the Netherlands, whereas students are considered immigrant, minority or non-Dutch if at least one parent is not born in the Netherlands (see Hartog, 2011). The share of students in our sample with a non-Dutch ethnicity is 0.55 and the percentage of students with a non-Dutch ethnicity varies per team from 20% to 90%.[5] We measure ethnic diversity dichotomously. This is consistent with previous empirical studies (Boisjoly et al., 2006; Carter et al., 2010; Hamilton et al., 2012; Hansen et al., 2006) as well as with our assignment of students to teams.[6] Students with a non-Dutch ethnicity are born abroad in 38% of the cases, with 82% of the fathers born abroad and 87% of the mothers. In 69% of the cases both the father and the mother are born abroad.

Table 4.1 shows that, based on the actual ethnicity of students (from the pretreatment survey), the students in our sample are from 53 different countries of origin with an average number of approximately 6 different countries of origin per team (panel A). For the purpose of presentation, the descriptive statistics in panel B are clustered by continent (see Parrotta et al., 2010). They indicate that ethnic variation among minority students is substantial. Panel C shows that there are virtually no clusters of ethnic minorities at the team level. The vast majority of ethnic minority students has no team members from the same country of origin (70%), whereas 19% of the ethnic minority students has only one team member from the same country of origin (the average number of students with a non-Dutch ethnicity per team is 6.8 at baseline). Hence, we can safely assume that the share of students with a non-Dutch ethnicity is a true indicator of the team's ethnic diversity.[7] Moreover, the average and range of the share of minorities are similar across fields of study (panel D). We will test for possible pretreatment differences more formally below.

[5]The average share of minorities in Dutch higher vocational schools is approximately 0.20, which is close to the population average (Hartog, 2011). The considerably higher share here reflects the international character of this program in the multicultural capital of Amsterdam.

[6]An alternative way of measuring ethnic diversity takes the country of (parental) origin into account and is expressed in terms of Blau's diversity index (1977), $1 - \sum_{i=1}^{N} p_i^2$. Here, p is the share of individuals in a (ethnic) category i and N the number of (ethnic) categories. Disadvantages of Blau's diversity index are that it: (i) implicitly assumes that each category is different from another *to the same extent* and (ii) turns out rather sensitive to the number of (ethnic) categories in the sample. The upside, though, may be that it takes into account that diversity is larger when individuals originate from all distinct countries than when they or their parents are all born in the same foreign country. However, our initial assignment of students to teams and, hence, the exogenous variation in teams' diversity is based on the dichotomous measure of ethnicity. As a consequence, we can only use that measure convincingly to study the causal effect of ethnic diversity on team performance.

[7]Only two teams form an exception: one accommodates six and the other four members from the same country of origin. Excluding these two teams from the sample does not change the main results. Robustness checks indicate that our findings also remain similar when we control for the number of different countries of origin.

Table 4.1. Descriptive statistics of ethnic diversity

A: Ethnicity (team level)	Mean	SD	Min	Max	
Share of minorities	0.55	0.16	0.20	0.90	
Number of different countries of origin	6.04	1.82	3.00	12.00	
B: Ethnicity (≈ continent)	Fraction				
Netherlands (native)	0.45				
North America / Oceania	0.01				
Central and South America	0.15				
Formerly Communist Countries	0.02				
Muslim Countries	0.17				
East Asia	0.06				
Asia	0.04				
Africa	0.03				
Other European Countries	0.07				
C: Ethnic minority students with number of team members from the same country of origin	Fraction				
0	0.70				
1	0.19				
2	0.05				
$\geqslant 3$	0.06				
D: Numbers of students and teams, and share of minorities by field of study	Students	Teams	Share of minorities (per team)		
			Average	Minimum	Maximum
Business management	240	18	0.52	0.20	0.79
Management	60	5	0.54	0.31	0.64
Trade management Asia	105	9	0.66	0.42	0.90
Business languages	118	11	0.50	0.22	0.77
Financial management	27	2	0.66	0.58	0.73
Total	550	45	0.55	0.20	0.90

Note: Ethnic minority students are from the following countries of origin: *North America / Oceania*, Australia, Canada, United States; *Central and South America*, Antilles, Argentina, Barbados, Brazil, Colombia, Costa Rica, Cuba, Dominican Republic, Guyana, Honduras, Jamaica, Peru, Surinam; *Formerly Communist Countries*, Russia and other East European countries; *Muslim Countries*, Afghanistan, Algeria, Egypt, Eritrea, Guinea, Indonesia, Iran, Iraq, Malaysia, Morocco, Nigeria, Pakistan, Palestine, Somalia, Tunisia, Turkey; *East Asia*, China, Hong Kong, Japan; *Asia*, other Asian countries not included in categories of East Asia and Muslim Countries (India, Laos, Philippines, Singapore, Vietnam); *Africa*, other African countries not included in category of Muslim Countries (Angola, Cameroon, Cape Verde, Gabon, Ghana, Kenya, South Africa); *Other European Countries*, other European countries not included in category of Formerly Communist Countries (Germany, Israel and other West European countries).

4.2.2 Other variables

Table 4.2 reports descriptive statistics of individual and team characteristics. The pretreatment survey administers background characteristics such as age and gender (see also Table 3.2 in Chapter 3). In addition, we measure scholastic achievements of students just before entering the college (indicated by 'grade point average'). All three surveys also include self-assessments of the knowledge that students have in three areas most relevant for successful entrepreneurship (see Karlan and Valdivia, 2011; Minniti and Bygrave, 2001). Hence, we can trace the individual development of these knowledge areas during the program as a proxy for learning. Knowledge levels at baseline are reported in Table 4.2.[8]

Table 4.2. Descriptive statistics of individual and team characteristics

	Scale	Mean	SD	Min	Max
Individual level					
Age	years	19.37	1.99	15.98	30.92
Gender (dummy = 1 if female)	0/1	0.44	0.50	0.00	1.00
Grade point average	1-10	6.46	0.24	6.05	7.23
Business knowledge	1-5	2.66	0.88	1.00	5.00
Entrepreneurship knowledge	1-5	2.71	0.98	1.00	5.00
Leadership knowledge	1-5	3.14	0.98	1.00	5.00
Team level					
Size (at baseline)	persons	12.22	2.09	8.00	16.00
Conflicts	1-5	2.23	0.59	1.00	3.67
Atmosphere	1-5	3.53	0.55	2.33	4.83

Business performance metrics are gathered or calculated from the companies' annual reports that we obtained from 43 out of 45 teams. We measure sales, profits and profits per share in euros. We also add a binary indicator of positive profits because students tend to view as the bottom line result whether they are able to satisfy their shareholders. Column (1) of Table 4.3 shows that average sales for all teams amount to 838 euros with a standard deviation of 707 euros. Profits are on average negative at -69 euros varying from a loss of 1016 euros to a profit of 477 euros. 22 teams make positive profits, while 21 teams run a loss. Profits per share vary between -15 and +15 euros.[9]

[8]Through standard batteries of questions we also obtain (non-tabulated) validated measures of 'softer' individual characteristics that are associated with entrepreneurship: the 'big five' factor model of personality structure (see Goldberg, 1990; Zhao and Seibert, 2006), and entrepreneurial skills and traits such as creativity, market awareness, networking, perseverance, need for achievement and risk aversion (see Chapter 3; Oosterbeek et al., 2010; Parker, 2009).

[9]The mean number of shares issued is 52 (s.d. 21.5), while the minimum and maximum numbers of shares sold are respectively 15 and 100. The number of shares is unrelated to teams' ethnic diversity.

Table 4.3. Descriptive statistics of business outcomes

		All	Ethnic diversity (0.45 and 0.65)			Ethnic diversity (0.40 and 0.60)		
			Low share$<$0.45	Moderate 0.45$\geq$share$\leq$0.65	High share$>$0.65	Low share$<$0.40	Moderate 0.40$\geq$share$\leq$0.60	High share$>$0.60
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sales (euros)	Mean	838	859	881	745	1183	826	735
	SD	707	634	891	349	622	895	468
	Min	0	130	0	265	593	130	0
	Max	4209	2226	4209	1267	2226	4209	1953
Profits (euros)	Mean	-69	-39	-146	41	-97	-78	-51
	SD	318	375	346	164	468	294	305
	Min	-1016	-1011	-1016	-247	-1011	-848	-1016
	Max	477	294	477	294	294	477	294
Positive profits (0/1)	Mean	0.51	0.70	0.29	0.75	0.67	0.37	0.61
	SD	0.51	0.48	0.46	0.45	0.52	0.50	0.50
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Profits per share (euros)	Mean	-0.51	0.42	-2.07	1.46	-1.84	-0.58	0.01
	SD	6.42	7.75	6.33	5.07	7.05	6.64	6.29
	Min	-15.48	-15.48	-12.87	-8.04	-15.48	-11.78	-12.87
	Max	15.64	15.64	11.36	12.79	3.67	15.64	12.79
N		43	10	21	12	6	19	18

Columns (2) through (7) of Table 4.3 present descriptive statistics of business outcomes for different degrees of ethnic diversity. We tabulate two pairs of cutoff levels around the average share of minorities in the sample of 0.55. Columns (2) through (4) with cutoffs at 45% and 65% of minorities indicate that teams of high ethnic diversity tend to perform better than teams of moderate ethnic diversity in terms of profits, the probability of profits being positive and profits per share. Performance differences are less pronounced in a comparison between teams of low and high ethnic diversity, although the descriptive statistics suggest that on average business outcomes are slightly higher for teams of high ethnic diversity.

Columns (5) through (7) show a similar pattern for more symmetric cutoff levels at a share of minorities of 0.40 and 0.60. However, we lack support among teams of low ethnic diversity due to the limited number of observations in the range below 40% of minorities (only 6 teams). In the remainder of this chapter, our main focus is therefore on teams with a moderate or high share of minorities (i.e., teams of moderate or high ethnic diversity).

Table 4.4. Randomization checks at the individual and team level

	Share of minorities			
A: Individual level	Native (Dutch) students		Ethnic minority students	
Age	0.805	(0.771)	-2.184**	(0.876)
Gender	-0.524**	(0.220)	0.067	(0.237)
Grade point average	-0.019	(0.089)	0.188	(0.117)
Business management	-0.284	(0.536)	-0.700	(0.463)
Management	-0.048	(0.383)	-0.222	(0.206)
Trade management Asia	0.459*	(0.265)	0.863*	(0.443)
Business languages	-0.363	(0.498)	-0.112	(0.377)
Financial management	0.235	(0.164)	0.171	(0.221)
B: Team level (average)	All students			
Age	-0.093	(0.473)		
Gender	-0.193	(0.166)		
Grade point average	-0.052	(0.058)		
Business management	-0.562	(0.432)		
Management	-0.075	(0.272)		
Trade management Asia	0.815**	(0.354)		
Business languages	-0.328	(0.408)		
Financial management	0.150	(0.148)		
Team size	1.579	(1.820)		

Note: In panel A each coefficient comes from a regression at the individual level of the row variable on the share of minorities, separately for native (Dutch) and ethnic minority students (robust standard errors in parentheses). In panel B each coefficient comes from a regression at the team level of the row variable on the share of minorities (bootstrapped standard errors in parentheses; 1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

4.2.3 Randomization

To examine whether students are randomly assigned to teams of different ethnic composition, we regress students' characteristics on the share of minorities in their team, separately for students of Dutch and non-Dutch ethnicity (see Table 4.4). The first two columns of panel A show that native (Dutch) students who are assigned to teams with many ethnic minority students are somewhat less likely to be female and more likely to study 'trade management Asia'. The last two columns of panel A indicate that ethnic minority students assigned to teams with a high share of minorities are somewhat more likely to be younger and to study 'trade management Asia'. In all other dimensions, the assignment of native (Dutch) and ethnic minority students is random.

Similarly, we examined at the team level whether (average) characteristics of students correlate with the ethnic composition of teams. Panel B shows that ethnic diversity is not systematically related to any of the (average) team characteristics except for the likelihood of studying 'trade management Asia'. Based on the analyses in this subsection all regressions in the next section include a dummy for this field of study.

4.3 Results

4.3.1 Main findings

Table 4.5 shows the relationship between the share of minorities in a team and four measures of business performance: sales, profits, the probability of profits being positive and profits per share. Note that a larger share of minorities implies a more ethnically diverse team due to the limited presence of ethnic clusters within these teams. In panels A and B performance measures are regressed on the share of minorities in the team and its square (using OLS, median and robust M-estimation regression).

The specifications in panel A testing for a linear effect of the share of minorities on business performance turn out insignificant. We also test polynomial specifications in panel B and these turn out being (largely) insignificant too, although the point estimates consistently suggest a u-shaped relationship between the share of minorities and business performance with the minimum at a share of about 0.55.

Panels C1 through C3 of Table 4.5 measure the effect of ethnic diversity on business performance using more flexible spline functions that allow for distinct slopes below (1st segment) and above (2nd segment) a certain share of minorities. Spline functions may be particularly informative in this case since the distribution of the share of minorities in our sample is asymmetric, i.e., teams with a relatively high share of minorities are actually heterogeneous in terms of ethnicity (instead of homogeneous and of one non-Dutch ethnicity). We employ different cutoff levels around the sample average of 55%

Table 4.5. Ethnic diversity and team performance

	Sales			Profits			Pos. profits	Profits per share		
	OLS	Median	Robust	OLS	Median	Robust	OLS	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: Linear										
% minorities	-597.8	-267.3	15.3	106.7	3.2	-71.4	0.153	3.779	-0.122	1.751
	(624.5)	(722.2)	(664.3)	(318.0)	(330.0)	(237.8)	(0.528)	(6.200)	(5.995)	(5.556)
R^2	*0.02*	*0.03*	*0.03*	*0.02*	*0.03*	*0.00*	*0.03*	*0.01*	*0.01*	*0.00*
B: Polynomial										
% minorities	-3798.5	-3780.9	-5009.1	-1741.8	-3018.7*	-2215.7	-4.622	-29.232	-40.341	-48.841
	(3968.5)	(4734.9)	(4322.8)	(2071.2)	(1621.2)	(1567.4)	(3.170)	(34.954)	(33.848)	(35.781)
% minorities^2	2942.8	3561.9	4306.0	1699.5	2664.7*	1942.6	4.391	30.350	36.524	45.266
	(3528.0)	(4040.7)	(3646.3)	(1671.1)	(1381.7)	(1316.9)	(2.693)	(28.902)	(29.662)	(29.096)
Minimum	*0.65*	*0.53*	*0.58*	*0.51*	*0.57*	*0.57*	*0.53*	*0.48*	*0.55*	*0.54*
R^2	*0.04*	*0.05*	*0.11*	*0.05*	*0.06*	*0.08*	*0.11*	*0.04*	*0.05*	*0.07*
C1: Spline (0.50)										
1st segment	-977.5	-903.9	-1709.2	-598.8	-955.1	-964.8	-1.816	-7.962	-18.541	-19.408
	(2308.9)	(2238.4)	(1938.0)	(1046.5)	(958.6)	(947.3)	(1.469)	(17.040)	(18.435)	(19.467)
2nd segment	-359.8	786.2	722.2	548.8	405.2	442.1	1.387*	11.137	11.402	13.080
	(1347.9)	(1072.7)	(834.8)	(398.8)	(445.5)	(401.3)	(0.811)	(8.285)	(10.001)	(8.892)
R^2	*0.02*	*0.05*	*0.08*	*0.05*	*0.05*	*0.08*	*0.11*	*0.03*	*0.05*	*0.07*
C2: Spline (0.55)										
1st segment	-1205.2	-903.9	-1102.3	-568.9	-955.1	-946.6*	-1.801*	-9.349	-17.288	-17.494
	(1404.8)	(1521.8)	(1451.2)	(766.7)	(647.4)	(559.9)	(1.088)	(13.563)	(10.851)	(12.553)
2nd segment	35.7	786.2	790.7	811.3*	1007.6**	757.9*	2.191**	17.471*	16.378*	19.601**
	(1199.3)	(1191.1)	(921.3)	(459.4)	(486.3)	(401.8)	(0.872)	(9.552)	(9.696)	(8.261)
R^2	*0.03*	*0.05*	*0.07*	*0.06*	*0.08*	*0.13*	*0.17*	*0.05*	*0.07*	*0.10*
C3: Spline (0.60)										
1st segment	-1200.1	-903.9	-814.2	-514.4	-718.5	-786.4*	-1.422*	-8.327	-14.429	-14.789
	(1012.2)	(1329.7)	(1191.7)	(630.3)	(497.0)	(424.9)	(0.854)	(10.805)	(9.575)	(10.222)
2nd segment	464.8	1733.7	1079.4	1202.3*	1257.0**	1042.3**	2.933**	25.136**	22.331*	26.773**
	(1236.0)	(1609.5)	(1155.7)	(632.2)	(607.1)	(517.3)	(1.143)	(11.410)	(12.322)	(10.950)
R^2	*0.03*	*0.05*	*0.06*	*0.08*	*0.08*	*0.14*	*0.18*	*0.07*	*0.08*	*0.12*

Note: Based on information from 43 teams. All specifications include a dummy for the field of study 'trade management Asia'. OLS, Median and Robust refer to the estimation method. Median and robust specifications for positive profits are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

of minorities, which also happens to be the share of minorities that minimizes business performance as suggested by the less flexible quadratic specifications. The results from these spline functions in panels C1 through C3 with a cutoff at respectively 50%, 55% or 60% of minorities are similar if we use other cutoffs such as at a share of minorities of 0.45 or 0.65 (not tabulated).

The columns of panel C1 report the relationship between ethnic diversity in the team and business performance above and below 50% of minorities. The coefficients are not significantly different from zero. Panel C2 shows the results for the cutoff at a share of minorities of 0.55. All point estimates for the share of minorities, given that this share is above 0.55, are positive (and except for sales) statistically significant. When the share of minorities is lower than or equal to 0.55, all coefficients are negative and in two cases significantly so. The point estimate in column (5) of panel C2 implies that profits increase by 100 euros (about one-third of a standard deviation) if the share of minorities is raised from 0.6 to 0.7. The columns of panel C3 show a similar pattern for the relationship between business performance and ethnic diversity above and below 60% of minorities: profits, the probability of profits being positive and profits per share only increase with the share of minorities in the segment above 60% of minorities.

As panels C1 through C3 of Table 4.5 show, these results are not driven by outliers since they are largely insensitive to using OLS, median or robust M-estimation regression. Table A1 in the appendix indicates (for various cutoff levels) that the results also remain similar if we estimate spline functions with three segments (i.e., with teams of low, moderate and high ethnic diversity). Here, business performance is lower for teams of low and moderate ethnic diversity relative to teams of high ethnic diversity, although the number of teams may slightly limit the precision of the estimates across different cutoffs.

The results presented in this subsection show that the relationship between team performance and ethnic diversity is flat or tends to decline down to a certain threshold level and starts increasing beyond this threshold level. The specific level of this threshold is around a share of minorities of 0.55. Hence, only if ethnic diversity is sufficiently substantial the net effect of ethnic diversity on team performance is positive. This finding is not at odds with the idea that the costs of coordination and communication offset the benefits of complementarities and (mutual) learning in homogeneous and moderately heterogeneous teams, while the benefits of sharing and exchanging relevant knowledge outweigh these costs in more heterogeneous teams.

4.3.2 Costs and benefits of ethnic diversity

In this subsection we explore whether our data, collected through the questionnaires, indeed suggest mechanisms that possibly drive the effect of ethnic diversity on team performance. Note that the mechanisms of interest are endogenous and that we lack exogenous variation to identify their causal impact (see Chapters 1 and 3). In what follows, we first consider the mechanisms that are associated with the costs of ethnic diversity and subsequently the mechanisms that are related to its benefits.

For communication and coordination, we expect that moderately heterogeneous teams incur higher costs of communication and coordination than teams that are homogeneous or heterogeneous in terms of ethnicity. We measure coordination and communication costs in terms of a team's atmosphere and personal conflicts (see Earley and Mosakowski, 2000; Richard et al., 2004). Teams' atmosphere is administered by asking students to rate the atmosphere within their team on a 5-points scale in the posttreatment survey. Likewise, conflicts in the team are surveyed by asking students to what extent there was conflict or disagreement between team members about personal matters (that are not task-related). More personal conflicts and a worse team atmosphere are expected to coincide with worse communication and coordination due to, for example, subgrouping along ethnic lines, distrust or stereotyping.

In line with Alesina and La Ferrara (2005), we find that homogeneous and moderately heterogeneous teams tend to experience less conflicts than heterogeneous teams (not tabulated). However, the data also indicate that teams' atmosphere and ethnic diversity are not significantly related. In sum, these results do not consistently suggest that the costs of communication and coordination might drive our main findings.

For the benefits, we expect a positive relationship between ethnic diversity and business performance due to more complementarities and (mutual) learning in ethnically diverse teams. We measure learning in terms of the development in three knowledge areas most relevant for successful entrepreneurship: business, entrepreneurship and leadership (see Karlan and Valdivia, 2011; Minniti and Bygrave, 2001). For each of these knowledge areas, the indicator of individual learning is the difference between the self-assessed level in the posttreatment and pretreatment questionnaire.

Table 4.6 reports results from least squares regressions (panels A and B) and spline functions (panels C1 through C3) of team-average learning in business, entrepreneurship and leadership knowledge on ethnic diversity. Again, we present results from using OLS, median and robust M-estimation regression. Panel A shows no significant linear effect of the share of minorities on learning, whereas panel B suggests a u-shaped relationship with the minimum at a share of approximately 0.55.

Spline functions in panels C1 through C3 with a cutoff at respectively 50%, 55% or 60% of minorities indicate that the relationship between learning and share of minorities is flat or declines down to a threshold level of about 0.55 and starts increasing beyond this threshold level. Hence, on average, members of ethnic heterogeneous teams learn more than members of homogeneous and moderately heterogeneous teams. Additional

Table 4.6. Learning and ethnic diversity

	Development in knowledge of								
	Business			Entrepreneurship			Leadership		
	OLS	Median	Robust	OLS	Median	Robust	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)	(10)
A: Linear									
% minorities	0.275	0.054	0.241	-0.090	0.061	-0.138	0.237	0.032	0.232
	(0.436)	(0.485)	(0.471)	(0.351)	(0.590)	(0.389)	(0.393)	(0.660)	(0.465)
R^2	*0.34*	*0.24*	*0.42*	*0.34*	*0.18*	*0.32*	*0.30*	*0.16*	*0.27*
B: Polynomial									
% minorities	-2.670	-1.874	-2.659	-3.455**	-3.835	-3.669	-3.789**	-3.825	-3.999**
	(1.970)	(2.385)	(2.476)	(1.727)	(2.889)	(2.540)	(1.799)	(2.617)	(1.903)
% minorities^2	2.690	1.891	2.514	3.088**	3.228	3.192	3.691**	3.468	3.891**
	(1.656)	(2.035)	(2.157)	(1.557)	(2.568)	(2.344)	(1.650)	(2.494)	(1.810)
Minimum	*0.50*	*0.50*	*0.53*	*0.56*	*0.59*	*0.57*	*0.51*	*0.55*	*0.51*
R^2	*0.38*	*0.27*	*0.47*	*0.39*	*0.25*	*0.37*	*0.38*	*0.24*	*0.33*
C1: Spline (0.50)									
1st segment	-1.067	-1.331	-0.897	-1.471	-2.057*	-1.741	-1.324*	-1.705	-1.607
	(0.919)	(1.182)	(1.244)	(0.913)	(1.088)	(1.138)	(0.760)	(1.029)	(1.090)
2nd segment	1.049*	0.867	0.779	0.753	0.858	0.752	1.191**	1.131	1.374**
	(0.623)	(0.656)	(0.622)	(0.530)	(0.745)	(0.624)	(0.549)	(0.852)	(0.686)
R^2	*0.38*	*0.26*	*0.46*	*0.39*	*0.26*	*0.37*	*0.37*	*0.26*	*0.36*
C2: Spline (0.55)									
1st segment	-0.786	-0.731	-0.881	-1.270**	-1.833**	-1.476**	-1.142*	-1.247	-1.287*
	(0.649)	(0.811)	(0.798)	(0.646)	(0.846)	(0.693)	(0.635)	(0.749)	(0.712)
2nd segment	1.321**	0.908	1.091	1.130*	1.219	1.114	1.662***	1.243	1.729**
	(0.656)	(0.757)	(0.685)	(0.599)	(0.876)	(0.711)	(0.626)	(0.869)	(0.729)
R^2	*0.38*	*0.28*	*0.49*	*0.41*	*0.27*	*0.41*	*0.41*	*0.28*	*0.37*
C3: Spline (0.60)									
1st segment	-0.493	-0.638	-0.731	-0.971*	-1.568**	-1.179**	-0.831	-1.175	-0.896
	(0.638)	(0.686)	(0.625)	(0.553)	(0.764)	(0.569)	(0.525)	(0.756)	(0.655)
2nd segment	1.585*	1.304	1.459	1.480*	1.540	1.499	2.133**	2.502*	2.129*
	(0.873)	(1.144)	(1.100)	(0.793)	(1.209)	(1.046)	(0.883)	(1.345)	(1.228)
R^2	*0.38*	*0.28*	*0.51*	*0.40*	*0.25*	*0.40*	*0.41*	*0.25*	*0.36*

Note: Based on information from 43 teams. All specifications include controls for team size, field of study 'trade management Asia' and team-average knowledge levels at baseline. OLS, Median and Robust refer to the estimation method. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

Table 4.7. Complementarities and diversity in relevant knowledge

	Complementarities			Diversity in knowledge of								
	Business/E'ship/Leadership			Business			Entrepreneurship			Leadership		
	OLS	Median	Robust	OLS	Median	Robust	OLS	Median	Robust	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A: Linear												
% minorities	0.690	1.397	1.181	0.129*	0.186	0.126	0.105	0.079	0.096	-0.059	-0.047	-0.049
	(0.533)	(0.872)	(0.920)	(0.076)	(0.127)	(0.094)	(0.109)	(0.160)	(0.132)	(0.071)	(0.105)	(0.084)
R^2	*0.28*	*0.22*	*0.41*	*0.54*	*0.32*	*0.46*	*0.17*	*0.03*	*0.14*	*0.23*	*0.13*	*0.19*
B: Polynomial												
% minorities	-3.776	-4.352	-2.763	-0.371	-0.639	-0.402	-0.766	-1.038	-0.786	-0.597	-0.272	-0.599
	(3.132)	(4.273)	(4.155)	(0.418)	(0.662)	(0.502)	(0.520)	(0.809)	(0.657)	(0.364)	(0.580)	(0.446)
% minorities^2	4.065	4.517	3.828	0.457	0.670	0.483	0.800*	0.978	0.811	0.493	0.211	0.503
	(2.697)	(3.784)	(3.742)	(0.371)	(0.574)	(0.433)	(0.475)	(0.733)	(0.605)	(0.323)	(0.518)	(0.402)
Minimum	*0.46*	*0.48*	*0.36*	*0.41*	*0.48*	*0.42*	*0.48*	*0.53*	*0.48*	*0.61*	*0.64*	*0.60*
R^2	*0.35*	*0.24*	*0.58*	*0.56*	*0.34*	*0.51*	*0.23*	*0.10*	*0.20*	*0.28*	*0.15*	*0.24*
C1: Spline (0.50)												
1st segment	-1.376	-1.680	-1.321	-0.025	0.040	-0.041	-0.239	-0.356	-0.246	-0.245	-0.156	-0.229
	(1.318)	(1.862)	(1.754)	(0.179)	(0.292)	(0.216)	(0.208)	(0.315)	(0.247)	(0.164)	(0.259)	(0.185)
2nd segment	1.878**	2.126*	2.144*	0.218*	0.234	0.224	0.315*	0.309	0.311	0.054	0.027	0.061
	(0.812)	(1.243)	(1.253)	(0.124)	(0.174)	(0.143)	(0.170)	(0.260)	(0.213)	(0.104)	(0.183)	(0.126)
R^2	*0.37*	*0.26*	*0.40*	*0.55*	*0.33*	*0.50*	*0.23*	*0.09*	*0.19*	*0.27*	*0.14*	*0.22*
C2: Spline (0.55)												
1st segment	-0.526	-0.054	-0.392	-0.016	-0.159	-0.033	-0.177	-0.272	-0.183	-0.201	-0.150	-0.189
	(1.053)	(1.437)	(1.325)	(0.148)	(0.232)	(0.182)	(0.162)	(0.244)	(0.191)	(0.135)	(0.185)	(0.144)
2nd segment	1.889*	1.960	2.167	0.273*	0.330	0.286	0.397**	0.464	0.391	0.086	0.031	0.099
	(1.000)	(1.649)	(1.451)	(0.150)	(0.217)	(0.188)	(0.196)	(0.319)	(0.258)	(0.122)	(0.218)	(0.159)
R^2	*0.34*	*0.24*	*0.37*	*0.56*	*0.34*	*0.51*	*0.24*	*0.11*	*0.19*	*0.27*	*0.14*	*0.22*
C3: Spline (0.60)												
1st segment	0.022	0.783	0.790	-0.012	-0.068	-0.032	-0.134	-0.167	-0.147	-0.172	-0.143	-0.167
	(0.929)	(1.312)	(1.246)	(0.119)	(0.190)	(0.158)	(0.139)	(0.216)	(0.173)	(0.111)	(0.167)	(0.124)
2nd segment	1.831	1.933	2.830	0.370*	0.401	0.397*	0.532**	0.507	0.534	0.141	0.054	0.157
	(1.335)	(1.932)	(1.843)	(0.190)	(0.279)	(0.229)	(0.254)	(0.395)	(0.329)	(0.165)	(0.271)	(0.197)
R^2	*0.31*	*0.22*	*0.57*	*0.56*	*0.35*	*0.52*	*0.26*	*0.11*	*0.21*	*0.27*	*0.14*	*0.23*

Note: Based on information from 43 teams. All specifications include controls for team size, field of study 'trade management Asia' and team-average knowledge levels at baseline. OLS, Median and Robust refer to the estimation method. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

regressions at the individual level, that we run separately for students of Dutch and non-Dutch ethnicity, show that the learning benefits of more ethnic diversity accrue to similar extents to both groups (not tabulated). Moreover, and probably as an explanatory factor of the higher learning levels in more diverse teams, the data show some evidence that ethnically diverse teams enter the entrepreneurship program at the start with a more diverse pool of relevant knowledge than less heterogeneous teams (see Lazear, 1999). Table 4.7 indicates that complementarities between team members and the coefficients of variation of business, entrepreneurship and leadership knowledge at baseline tend to be larger in ethnically diverse teams.[10]

All in all, based on these results we cannot reject the idea that ethnic diversity benefits (mutual) learning and heterogeneous knowledge, possibly leading to better team performance. This finding is partly consistent with the theoretical ideas formulated in the introduction of this chapter. Ethnically diverse teams tend to have a more diverse pool of relevant knowledge and (possibly based on this) experience more learning and achieve better results. However, we do not find support for the idea that moderately heterogeneous teams incur higher costs of coordination and communication.

4.4 Discussion and conclusion

This chapter shows evidence of a positive impact of ethnic diversity on team performance, although only starting at a certain turning point at which at least half of the team is ethnically diverse. Before this turning point the relationship is flat or slightly negative. Hence, only if ethnic diversity is sufficiently substantial the net effect of ethnic diversity on team performance is positive. In line with theoretical predictions (Lazear, 1999), our data suggest that ethnic heterogeneous teams benefit from a more diverse pool of relevant knowledge facilitating (mutual) learning.

Our study is motivated by the fact that many decisions in organizations are nowadays assigned to teams (Hamilton et al., 2003) that become increasingly diverse due to the changing composition of Western populations (e.g., Ozgen et al., 2011b). One of the most salient and relevant dimensions of team heterogeneity is ethnicity (Alesina and La Ferrara, 2005). Until today, however, studies analyzing the causal effect of ethnic diversity on team performance in the longer run have been scarce.

[10]In Table 4.7 a diverse pool of relevant knowledge at the start of the program is operationalized by: (i) complementarities between the self-assessed knowledge that team members have in business, entrepreneurship and leadership, and (ii) the coefficients of variation of business, entrepreneurship and leadership knowledge in teams at baseline. Complementarities are constructed by first standardizing all three knowledge dimensions, subsequently computing the teams' maximum for each knowledge dimension, and then determining the teams' minimum of the maximums of all three knowledge dimensions. Supposedly, if students of different ethnicity complement each others knowledge, these minimums are higher in ethnically diverse teams.

Several limitations pertain to this study. There are discrepancies between the business teams in our study and teams in business practice. Individuals in our teams are relatively young, lack serious labor market experience and some of the teams have unprecedented high degrees of ethnic diversity. These characteristics might, to some extent, limit the external validity of our study. Moreover, although advantageous for the internal validity of our study, the random composition of teams is probably not representative of common practice in business. Finally, our experimental design does not allow for a causal interpretation of mechanisms such as (mutual) learning that lead to higher performance of ethnic heterogeneous teams.

Nevertheless, teams' substantial and genuinely joint task with strong incentives to maximize shareholder value of a real company in which team members have time to establish roles and observations of other members closely resembles the functioning of teams in business practice. Given the upcoming increase of the share of minorities in the labor force it is likely that any team will become more and more diverse in terms of ethnicity. Our study might provide a realistic preview of the impact that a high degree of ethnic diversity may have on the performance of teams.

Appendix Table A1. Ethnic diversity and team performance (spline functions with three segments)

	Sales			Profits			Pos. profits	Profits per share		
	OLS	Median	Robust	OLS	Median	Robust	OLS	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
C1: Spline (0.50 and 0.60)										
1st segment	-72.3	-903.9	-1475.2	-25.5	19.5	-298.3	-0.566	5.519	-1.568	-1.451
(N=13)	(3383.4)	(2832.8)	(2542.9)	(1305.0)	(1091.9)	(1186.7)	(1.553)	(20.586)	(21.003)	(21.464)
2nd segment	-3855.7	-1702.8	379.6	-1665.6	-2380.6	-1827.1	-3.438	-40.931	-44.992	-45.332
(N=12)	(5694.8)	(4160.3)	(3194.1)	(1924.9)	(1993.9)	(1789.0)	(2.537)	(34.724)	(41.217)	(38.860)
3rd segment	1121.1	1863.9	832.5	1486.8**	1533.8*	1272.7*	3.431***	33.194**	30.027*	33.510**
(N=18)	(1511.8)	(1730.1)	(1324.0)	(750.4)	(836.2)	(743.1)	(1.329)	(13.609)	(17.104)	(15.465)
R^2	*0.05*	*0.05*	*0.07*	*0.09*	*0.09*	*0.16*	*0.20*	*0.09*	*0.09*	*0.14*
C2: Spline (0.45 and 0.65)										
1st segment	-1982.7	-4112.3	-3445.7	-70.7	-729.5	-546.6	-0.833	9.147	-14.834	-3.369
(N=10)	(3386.2)	(3331.4)	(3215.0)	(1601.9)	(1746.8)	(1919.0)	(2.146)	(28.721)	(31.870)	(37.745)
2nd segment	-441.6	-52.1	660.5	-362.8	-376.4	-463.8	-0.804	-13.556	-6.730	-14.604
(N=21)	(2170.7)	(1717.5)	(1429.9)	(782.4)	(971.8)	(727.6)	(1.413)	(17.582)	(22.359)	(18.522)
3rd segment	593.0	1953.7	869.7	1233.6	1519.5	1091.6	3.130*	32.469*	25.126	32.904
(N=12)	(1832.0)	(1960.7)	(1948.5)	(816.3)	(1036.6)	(952.3)	(1.753)	(16.937)	(24.939)	(21.982)
R^2	*0.04*	*0.06*	*0.14*	*0.05*	*0.06*	*0.08*	*0.13*	*0.06*	*0.05*	*0.08*
C3: Spline (0.40 and 0.60)										
1st segment	-3825.4	-8217.8	-5895.6	542.9	1449.7	-107.4	1.138	26.214	34.250	9.394
(N=6)	(4174.3)	(4985.2)	(5886.8)	(2788.9)	(3150.5)	(4286.7)	(3.301)	(48.628)	(53.268)	(77.093)
2nd segment	179.6	749.1	662.8	-1070.0	-1424.8*	-1119.5	-2.768**	-26.479	-32.143*	-28.274
(N=19)	(1455.4)	(1535.9)	(1291.8)	(828.7)	(807.7)	(680.6)	(1.213)	(17.121)	(16.135)	(18.670)
3rd segment	-91.0	786.2	487.1	1426.1**	1340.6*	1187.0**	3.475***	32.448***	27.426**	31.785**
(N=18)	(1172.0)	(1482.4)	(1219.8)	(685.2)	(693.3)	(584.7)	(1.169)	(12.559)	(13.511)	(13.311)
R^2	*0.05*	*0.08*	*0.22*	*0.10*	*0.10*	*0.15*	*0.22*	*0.11*	*0.10*	*0.14*

Note: Based on information from 43 teams. All specifications include a dummy for the field of study 'trade management Asia'. OLS, Median and Robust refer to the estimation method. Median and robust specifications for positive profits are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

Chapter 5

Ability dispersion and team performance

This chapter is based on Hoogendoorn, S., S. Parker, and M. van Praag (2012), Ability dispersion and team performance: A field experiment, Tinbergen Institute Discussion Paper No. 2012-130/7.

5.1 Introduction

Few topics in economics are not associated with cognitive ability. 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). Ability considerations carry over naturally from individual settings to those involving teams of individuals: studies of team performance invariably emphasize the importance of individual team member abilities (e.g., Stewart, 2006).

Yet, despite it being widely believed that (cognitive) abilities of members affect overall team performance, the precise impact of ability dispersion on team performance remains poorly understood (Hamilton et al., 2012). Specifically, we lack evidence about whether ability diversity of team members is or is not conducive to the performance of teams. Such evidence is potentially useful for managers seeking to select members of internal work teams, as well as being of interest in its own right.

This chapter examines the effect of the dispersion of cognitive ability (hereafter, just 'ability') on the performance of business teams in a field experiment. Our experiment randomizes 573 students into 49 teams conditional on their measured cognitive abilities. We ensure a relatively large exogenous variation in ability dispersion between teams to help probe non-linearities in the relationship between ability dispersion and team performance.

We believe there are two principal advantages of our empirical design. First, a field experiment can establish a causal relationship between team composition and performance, in contrast to regression-based approaches in which members are free to self-select into and out of teams (Hansen et al., 2006). Second, relative to prior empirical studies which have analyzed settings involving laboratory experiments (Woolley et al., 2010), unskilled work tasks (e.g., Hamilton et al., 2003), or competitive sports (Kahn, 2000), our field experiment closely resembles the functioning of teams co-operating on a complicated real business project. The case of business (management) teams is worth understanding well. These teams often exert considerable influence on the performance of public and private organizations since they take decisions of strategic and operational importance. Moreover, the tasks of these teams are often complex and broad in scope, entailing the sustained application of members' cognitive abilities. This is precisely the context of the field experiment conducted in this chapter.

Our findings can be summarized as follows. Team performance as measured in terms of sales, profits and profits per share first increases at low levels of ability dispersion up to a maximum before decreasing at higher levels of dispersion. Controlling for the average cognitive ability of teams, performance is maximized at a coefficient of variation in cognitive ability of approximately 0.25 (the sample average is 0.22). Teams of moderate ability dispersion also experience fewer dismissals during the entrepreneurship program, although this does not chiefly explain why those teams perform better.

Prior theorizing turns out to be unable to explain these results. Previous theories have highlighted benefits to, or costs from, diversity - predicting linear relationships between member ability diversity and team performance. Thus, if the inputs of members are complementary in team production, performance is maximized by matching individuals of similar abilities (Kremer, 1993). However, if members' inputs are substitutable, heterogeneous teams in terms of ability maximize performance (Prat, 2002).[1] These arguments have been applied to explain performance in team sports, for example, Gould and Winter (2009) in the case of Major League Baseball, and Franck and Nüesch (2010) in the case of German professional soccer. Other reasons why teams might benefit from heterogeneity in ability among their members include the imposition of valuable high team production norms by a few powerful able members and learning by less able from more able team members (Hamilton et al., 2003).[2]

[1] In a model of conjunctive team production the lowest-ability member (i.e., the weakest link) in the extreme case determines team performance, whereas a model of disjunctive team production implies that ultimately the highest-ability member (i.e., the superstar) determines the performance of teams (see also Hong and Page, 2001, 2004; Iranzo et al., 2008; Rosen, 1981).

[2] In line with the predictions of their model, Hamilton et al. (2003) find that teams' ability dispersion positively affects their output using high-frequency productivity data from a garment plant.

In order to understand our results, we develop a novel model in which intra-team ability dispersion has both beneficial and detrimental effects. Low levels of ability dispersion in the model are associated with smaller pools of knowledge and, hence, team outcomes of limited scope and value. Greater dispersion is associated with larger pools of knowledge that produce more valuable team output, but this comes at an increasing cost. Team members can choose to shirk rather than provide effort, creating a free-riding incentive. Teams discourage shirking by monitoring their members; but the success of the monitoring technology is stochastic, decreasing with team member diversity.[3] At sufficiently high levels of ability dispersion, the probability of being caught and dismissed for shirking declines so much that shirking reduces team performance. Hence, moderate levels of ability dispersion in a team are associated with maximal team performance, i.e., the relationship between team performance and ability dispersion exhibits an inverse u-shaped pattern.

The remainder of this chapter is organized as follows. Section 5.2 presents the theoretical model. Section 5.3 describes the data of our field experiment. Section 5.4 presents the empirical findings. Section 5.5 discusses and concludes.

5.2 Model

5.2.1 Set up

We commence by analyzing a two-member team, in which both members create value by combining their knowledge. Members have potentially different cognitive abilities, $x_i \in X$ and $x_j \in X$, where X is the set of cognitive abilities in the general population. We have $\min\{X\} = x_L$ and $\max\{X\} = x_U > x_L$. Denote the average cognitive ability of the team by $\bar{x} = (x_i + x_j)/2$ and the difference between the two cognitive abilities by $d = |x_j - x_i|$: $d \in [0, x_U - x_L]$. There is a large number of teams in the population, whose total mass is unity. Decisions of one team do not affect other teams. Different teams may have different values of $(\bar{x}, d)$; our analysis will focus on only a single team.

People of given cognitive ability also have a given set of knowledge, denoted by $\{k\} \subset K$, where K is the set of knowledge in the general population. We assume that cognitive ability x and knowledge set $\{k\}$ are related by a *correspondence*, whereby each element x of X is related via a map g to a given subset $P(K)$ of K. Formally, $g : X \to P(K)$, where $P(K)$ is known as the *power set* of K. To crystallize ideas in what follows, we will work with the particular case where the power set associated with x is $\{x \pm \frac{1}{2}\Delta\}$, where $0 < \Delta < x_U - x_L$. Hence, $K = [x_L - \Delta/2, x_U + \Delta/2]$. In this

[3] A related deterrence mechanism is peer monitoring. For an analysis of peer effects in the workplace, see Falk and Ichino (2006), Kandel and Lazear (1992), and Mas and Moretti (2009).

case, a unique and equally sized set of knowledge, spanning Δ in size, is associated with each unique level of cognitive ability. For now, Δ will be taken to be invariant to x. Later on, we will relax this assumption and allow $\Delta = \Delta(x)$, with $\Delta'(x) > 0$ and $\Delta''(x) \geq 1$, to encompass the possibility that more able people have larger knowledge sets than less able people.

Note that team members of low and high cognitive ability both have productive roles in this set-up. That is because both members possess unique knowledge. For example, a very able team member might possess detailed scientific knowledge about an invention, which the less able team member may lack. Yet, the less able member might possess knowledge about, e.g., salesmanship or market conditions, which the more able member lacks. Team members pool their knowledge: team value, v, therefore depends on the union of their (non-overlapping) knowledge sets. Thus, the value of a team is greater than that of a single team member, as long as $x_j \neq x_i$, i.e., as long as $d > 0$.

Baumol and Strom (2010) describe how Matthew Boulton's knowledge of eighteenth century industrial market needs complemented James Watt's technical knowledge of steam engines and crank technology, forging a partnership that pioneered and disseminated the rotary-motion engines which powered the Industrial Revolution. As this example shows, the union of disparate knowledge generates value without either partner needing to acquire the knowledge possessed by the other (see also Lazear, 1999).

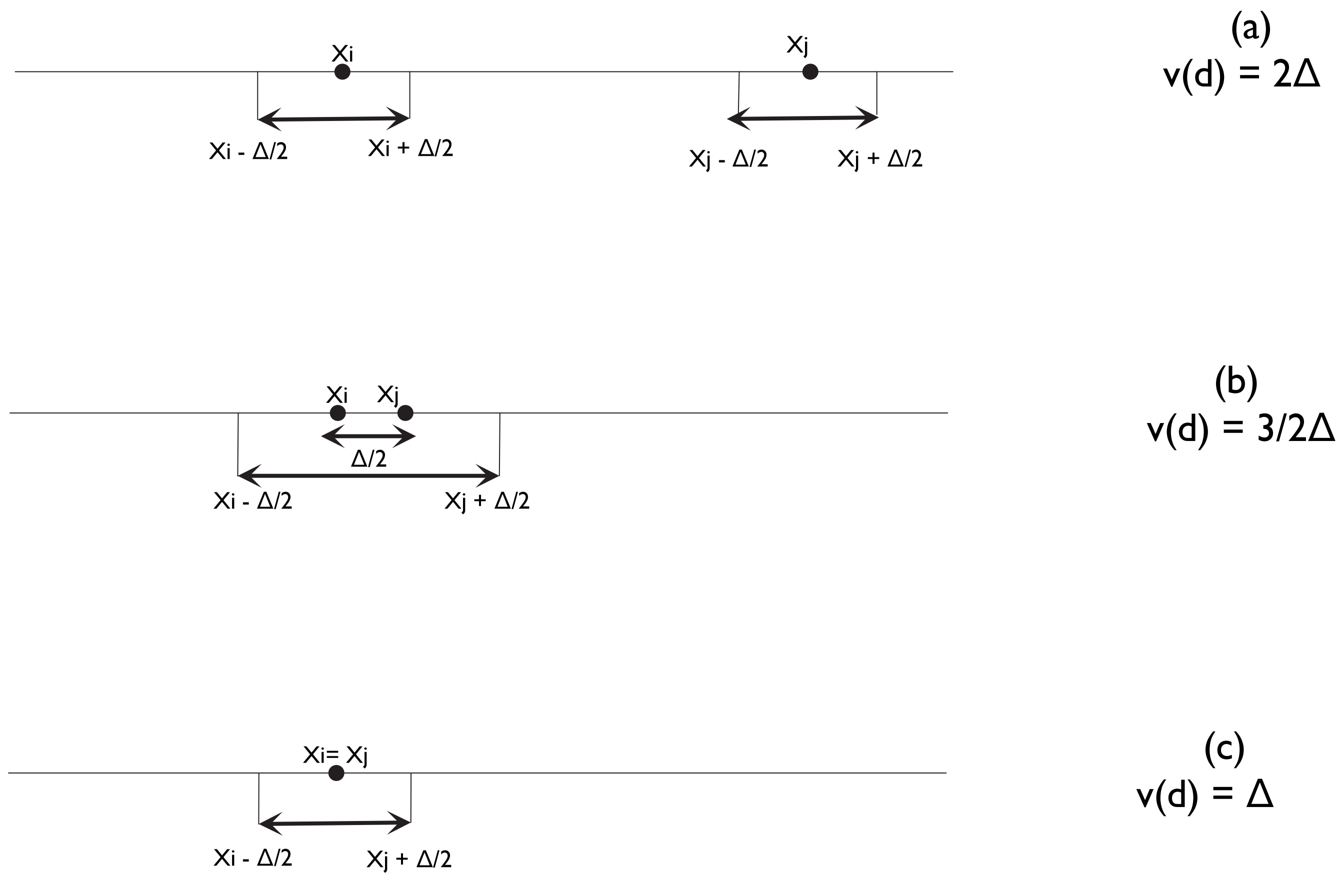
If x_i and x_j are very similar, knowledge spans will overlap and the gains from being in a team are modest. Hence, team value is greater the larger the ability difference d as long as $d < \Delta$. Once $d \geq \Delta$, team value is no longer increasing in d since maximal knowledge pooling has occurred. Hence, the team's value is increasing in v, where

$$v = v(d) = \begin{cases} \Delta + d & \text{for } d < \Delta \\ 2\Delta & \text{for } d \geq \Delta \end{cases} \tag{5.1}$$

See Figure 5.1 for an illustration of three different cases corresponding to (5.1). Note that none of the analysis or results that follow depend on the discontinuity of $v(d)$ in (5.1): a smooth function for $v(d)$ which exhibits diminishing marginal returns to d would generate the same qualitative results.

Total team value depends not only on $v(d)$ but also on the efforts of both team members. Effort is privately costly for each member but is non-contractible and observable only by the member who exerts it: it cannot be credibly communicated or signaled to the co-member (e.g., Laffont and Martimort, 2002). Each of the member efforts are inputs into the production process. Each effort, denoted by e_i and e_j for the two members, either takes a value of one (full effort, which incurs private and non-publicly-observed idiosyncratic effort costs of $c_i > 0$ and $c_j > 0$ respectively, where $c_i \neq c_j$ in general) or zero ('shirking': no effort, so no effort costs). Effort costs are

Figure 5.1. Illustration of three different cases corresponding to (5.1)



also private information, which cannot be credibly communicated or signaled to the co-member or a third party. The density function of effort costs in the population is $\gamma(c) > 0$ for $c \geq 0$; the cumulative distribution function is $\Gamma(c) = \int_0^c \gamma(\chi)\, d\chi \in [0, 1]$.

Each team member chooses their effort given expectations (derived below) of the other member's effort. Total gross team value is

$$\tilde{V} = v(d)[f(e_i, e_j) + \tilde{\omega}] \tag{5.2}$$

where f is an increasing function of both arguments. Without any important loss of generality, we will assume that $f(e_i, e_j) = e_i + e_j$. Hence, $f(0, 0) = 0$: zero effort by both members yields zero expected team value. The separability of the production function helps keep the decision-making separable at the individual level, which will enable us to study a game in pure strategies with a unique equilibrium. Let ω denote the outcome of a mean-zero random variable $\tilde{\omega}$ (e.g., this could capture stochastic demand for the team's output). Only the joint return $e_i + e_j + \omega$ is observed. So while a member knows her own effort with certainty, she cannot infer the other effort once V is observed, since ω is also unobserved.

This is important because it means that monitoring is the only way that members can measure the effort of co-members. At the outset, both members agree on the following contract: (a) realized team value is equally shared at the end of the period if neither member is dismissed, and (b) any member who is discovered via monitoring to have supplied zero effort is dismissed from the team. They receive zero output and pay a penalty of $\sigma > 0$. The remaining member takes the remaining output in its entirety. The reason for (a) is that costs and efforts are unobserved and output is a joint product: hence, an alternative *ex ante* compensation scheme based on effort-related equity shares is not feasible. For (b), dismissal in response to zero effort by the other member is optimal for the effort-providing member while providing an *ex ante* incentive for both members to exert costly effort. The penalty σ paid by a dismissed member could be lost capital or lost reputation, which detracts from their future expected economic prospects. The probability of dismissal will be derived endogenously below.

With probability $p(d)$, monitoring reveals to a member any shirking by her co-member. Here, p is assumed to be a smooth and continuous function. Detection of shirking is assumed to be more difficult in teams where there is greater distance d between the members, reflecting the difficulty of disentangling effort from the productivity of that effort.[4] Hence, $p'(d) < 0$, with $p(0) = 1$, $p(\infty) = 0$ and $p''(d) > 0$.

[4] Thus, taking a hypothetical extension of the Boulton and Watt example, Watt might be unsure about whether any initial lack of success by Boulton in selling engines is attributable to limited sales effort by Boulton or rather full but ineffective sales effort. Likewise, Boulton might be unsure whether any initial lack of success by Watt in building a new engine is attributable to limited technical effort by Watt or rather full but ineffective technical effort. Our assumption of $p'(d) < 0$ is consistent with

We assume that the monitoring technology never falsely indicates shirking when effort was actually supplied. Monitoring is taken to be effortless and costless without loss of generality.

We do not study the determinants of team formation, taking (in accordance with our empirical data) the composition of teams to be exogenous. It is worth briefly underlining the importance of assuming exogenous team composition. We want to predict how individual and team efforts, performance and dismissals vary with different amounts of team diversity. If instead we allowed for endogenous self-selection of diverse individuals into teams, we would only be able to analyze these relationships for that subset of team diversities associated with some given (assumed) self-selection process. We are interested in analyzing what could happen under a range of alternative sorting arrangements. Furthermore, the assumption of exogenous team assignment fits directly with our experimental design, which can therefore provide causal evidence about the consequences of a range of different team diversities.

5.2.2 Payoffs and optimal effort choices

Both members know d, but are symmetrically uninformed about the private effort costs faced by their co-members and, hence, the effort that their co-members will supply. They each believe with probability θ (derived below) that their co-member will exert high-effort and with probability $1-\theta$ that they will shirk. Expected payoffs under low and high effort by member j are

$$E(R_j|e_j=0) \quad = \quad \theta v(d)(1-p(d))/2 - p(d)\sigma \tag{5.3}$$

$$E(R_j|e_j=1) \quad = \quad v(d)\left[\theta + \frac{(1-\theta)(1-p(d))}{2} + (1-\theta)p(d)\right] - c_j \tag{5.4}$$

Evidently, j's optimal effort choice is given by

$$e_j^* = \begin{cases} 0 & \text{if } E(R_j|e_j=0) \geq E(R_j|e_j=1) \\ 1 & \text{otherwise} \end{cases} \tag{5.5}$$

After rearranging and collecting terms, the condition for $e_j^*=1$ is

$$c_j < v(d)(1+p(d))/2 + p(d)\sigma \tag{5.6}$$

We assume that the equilibrium strategy is the same for both members, i.e., an analogous expression to (5.6) exists for i (replace c_j by c_i on the LHS). Hence, we have a symmetric problem.

the idea that had Boulton and Watt been more dissimilar, their difficulties of disentangling lack of effort from ineffective deployment of effort would have been even greater.

Member j knows her own c_j but not c_i; i knows c_i but not c_j. While j does not know c_i she does know the distribution of types, so her subjective probability that the inequality in (5.6) holds is

$$\theta = \theta\big(v(d), p(d), \sigma\big) = \Gamma\left[v(d)(1+p(d))/2 + p(d)\sigma\right] \tag{5.7}$$

Differentiate (5.7) to obtain:

$$\frac{\partial \theta}{\partial v(d)} = \frac{\gamma(\cdot)(1+p(d))}{2} > 0 \tag{5.8}$$

$$\frac{\partial \theta}{\partial p(d)} = \gamma(\cdot)\left(\frac{v(d)}{2} + \sigma\right) > 0 \tag{5.9}$$

The signs of these derivatives both make intuitive sense: a greater return from effort increases the probability that effort is exerted, as does a greater probability of being caught if one shirks. Note that θ is fixed for a given team (i.e., for a given d), though its value varies across teams with different values of d. Because the researcher does not observe each private effort choice either, we can state our first proposition in terms of expected performance across different teams.

Proposition 1. *Provided the effectiveness of monitoring does not decline too rapidly as members become infinitesimally diverse, i.e., provided that*

$$|p'(0)| < \frac{2[\Gamma(\Delta+\sigma) + \Delta\gamma(\Delta+\sigma)]}{\gamma(\Delta+\sigma)(\sigma+\Delta/2)} \tag{5.10}$$

then expected team performance exhibits an inverse u-shaped relationship with the diversity of team member ability.

Proof. Both members provide effort with probability θ^2, while only one member provides effort with probability $\binom{2}{1}\theta(1-\theta)$. Hence, expected team performance and its derivative with respect to d, are

$$E(\tilde{V}) = 2v(d)\big(\theta^2 + \theta(1-\theta)\big) = 2v(d)\theta$$

$$\frac{\partial E(\tilde{V})}{\partial d} = 2\theta v'(d) + 2v(d)\left(\frac{\partial \theta}{\partial v(d)}v'(d) + \frac{\partial \theta}{\partial p(d)}p'(d)\right) \tag{5.11}$$

As shown above, the two derivatives (5.8) and (5.9) which appear in (5.11) in large brackets are both positive. Also $v'(d)$ is positive for $d < \Delta$ and zero for $d > \Delta$, while $p'(d) < 0\ \forall d$. Hence, the overall derivative (5.11) is certainly positive at $d = 0$ if the condition in the body of the proposition holds, and remains positive as long as $d < \Delta$

since $p''(d) > 0$. But for $d > \Delta$, $v'(d) = 0$ so by inspection (5.11) turns negative. Hence, expected team performance $E(\tilde{V})$ is an inverse u-shaped function of d.

The intuition for Proposition 1 is as follows. Low levels of team diversity ($0 \leq d < \Delta$) are associated with smaller pools of knowledge and, hence, team outcomes of limited scope and value. Higher levels of dispersion are associated with larger pools of knowledge that produce more valuable team output. However, there is a limit to this benefit of greater diversity; at sufficiently high levels of dispersion the probability of being caught and dismissed for shirking induces such low member effort that teams lose productivity and team performance worsens.

The role of the condition in the body of Proposition 1 has a straightforward interpretation. If monitoring effectiveness declines very rapidly as members become marginally different from each other, then effort declines so rapidly that the ensuing decline in performance dominates any positive effect from greater knowledge pooling. Thus, if the condition in the proposition does not hold, team performance will be strictly declining in d.

5.2.3 Dismissals

Next we ask which teams are most prone to dismissals of one member. In what follows, we will not pay attention to cases where both members end up dismissing each other (a case referred to as 'team collapse'). Indeed, there are no team collapses in the data to motivate such an analysis here. We will instead focus on cases where only one member is dismissed.

A dismissal occurs if one member shirks and is caught while the co-member either does not shirk (or does and is not caught; if both members are caught shirking there is a team collapse instead). There are two ways the event 'one member caught shirking' can occur, so the probability of a dismissal, $\Psi(d, \theta)$, is

$$\Psi(d, \theta) = 2p(d)(1-\theta)[\theta + (1-\theta)(1-p(d))] \tag{5.12}$$

The derivative with respect to d is

$$\frac{\partial \Psi(d, \theta)}{\partial d} = 2(1-\theta)\Upsilon(d)p'(d) - 2p(d)\Upsilon(d)\left(\frac{\partial \theta}{\partial v(d)}v'(d) + \frac{\partial \theta}{\partial p(d)}p'(d)\right) \tag{5.13}$$

where $\Upsilon(d) = 1 - 2p(d)(1-\theta)$. If (as is supported by the evidence in our study) less than half of all teams contain at least one detected shirker, then $p(d)(1-\theta) < 0.5$, whence $\Upsilon(d) > 0 \ \forall d$. A sufficient condition for this to hold is $2\Gamma(\Delta + \sigma) > 1$, which we will assume for the next proposition.

Proposition 2. *A sufficient condition for the incidence of dismissals to be a strictly decreasing function of diversity, d, is*

$$p(d)\frac{\partial\theta}{\partial p(d)} < 1-\theta \quad \forall d \tag{5.14}$$

If the inequality in (5.14) is reversed, and if in addition

$$|p'(0)| < \frac{\gamma(\Delta+\sigma)}{1-\Gamma(\Delta+\sigma)-\gamma(\Delta+\sigma)(\sigma+\Delta/2)} \tag{5.15}$$

then the incidence of dismissals is a u-shaped function of d.

Proof. Rearrange (5.13) to obtain:

$$\frac{\partial\Psi(d,\theta)}{\partial d} = 2\Upsilon(d)p'(d)\left[1-\theta-p(d)\frac{\partial\theta}{\partial p(d)}\right] - 2p(d)\Upsilon(d)\frac{\partial\theta}{\partial v(d)}v'(d) \tag{5.16}$$

If (5.14) holds, then (5.16) is strictly negative. If (5.14) does not hold while (5.15) does, then the final negative term of (5.16) dominates at $d = 0$, implying Ψ initially declines in d. For $d > \Delta$ we have $v'(d) = 0$ and the first term of (5.16) remains. This is positive if the inequality in (5.14) is reversed, proving the result.

The two possible dismissal profiles outlined in Proposition 2 depend on the sensitivity of effort to the probability of detecting shirkers. If effort is relatively insensitive to the probability that shirkers are detected (i.e., if (5.14) holds) then the likelihood of dismissals decreases as knowledge pooling generates benefits from effort (for low to moderate dispersion) and as the effectiveness at catching shirkers decreases (at high diversity). On the other hand, if effort is highly sensitive to the probability that shirkers are detected (i.e., if the inequality in (5.14) is reversed), and if the probability of detecting shirkers does not decrease too rapidly in response to a marginal increase in dispersion (condition (5.15)), then teams with low diversity supply more effort in response to benefits from knowledge pooling - leading to fewer dismissals initially. However, in more diverse teams where the marginal benefits of pooling have attenuated, members supply so much less effort that, even though the probability of being detected has decreased, there are more culprits to catch and hence the incidence of dismissals rises.

5.2.4 A simple extension

The model can be extended to treat the case in which cognitive ability affects the size of knowledge sets. This possibility introduces a novel prediction, summarized in

Proposition 3 below. To model this extension, suppose $\Delta'(x) > 0$ and $\Delta''(x) \geq 1$, so if $x_j > x_i$ then $\Delta(x_j) > \Delta(x_i)$. Then the following equation replaces (5.1):

$$v(d) = v(d, x_i, x_j) = \begin{cases} \Delta(x_j) & \text{for } d < [\Delta(x_j) - \Delta(x_i)]/2 \\ d + [\Delta(x_j) + \Delta(x_i)]/2 & \text{for } [\Delta(x_j) - \Delta(x_i)]/2 \leq d < [\Delta(x_j) + \Delta(x_i)]/2 \\ \Delta(x_j) + \Delta(x_i) & \text{for } d \geq [\Delta(x_j) + \Delta(x_i)]/2 \end{cases} \quad (5.17)$$

Proposition 3. *If knowledge sets and cognitive ability are positively related, and the condition (5.10) of Proposition 1 holds, then*

(a) Expected team performance exhibits first a declining and then an inverse u-shaped relationship with the diversity of team member ability, and

(b) Expected team performance is positively related to average team ability.

Note that there are two principal differences between Proposition 3 and Proposition 1. First, the flat initial section of the $v(d)$ function in (5.17) gives rise to an initial declining segment of the effort and performance relationships with respect to d, before the inverse u-segment emerges. A testable implication of this prediction is that a third-order polynomial function is needed to represent the performance-dispersion relationship if knowledge breadth really does increase with cognitive ability. In contrast, a third-order term will be insignificantly different from zero if knowledge breadth is invariant to cognitive ability, as assumed to be the case in Proposition 1.

Second, Proposition 3 predicts the importance of including an additional explanatory variable in a model of team performance if knowledge breadth increases with cognitive ability: namely, the average cognitive ability of team members. In contrast, this explanatory variable will be insignificantly different from zero if knowledge breadth is invariant to cognitive ability, as assumed to be the case in Proposition 1. Like the significance of the higher order term, this too is an easily testable prediction.

5.3 Data

We accessed various data sources to collect information about individuals and teams. One week before the start of the entrepreneurship program students took the 20-minute timed version of Raven's advanced progressive matrices test and filled out a pretreatment questionnaire that mainly covered their background characteristics (response rate: 89%). Simultaneously, we received administrative data to assist us in assigning students to teams. At the end of the program, students filled out a posttreatment questionnaire that queried team characteristics and processes (response rate: 68%).[5] We then also obtained the approved annual reports, which contain information about the business performance of teams. The data that we collected were used to construct exogenous variation in cognitive ability across teams (see Chapter 2), to test the predictions of our model (see Section 5.4) and to assess whether the assignment of students to teams was random conditional on their cognitive ability (discussed in this section).

5.3.1 Cognitive ability

Raven's advanced progressive matrices are extensively used to differentiate between people of higher cognitive ability (Bors and Stokes, 1998; Mills et al., 1993; Raven et al., 1993). The test requires subjects to select the missing figure out of eight possibilities that completes a logical pattern (see Figure A1 in the appendix for an example). Patterns become increasingly difficult as subjects progress. Over the past decades, Raven's advanced progressive matrices have been shown to associate with cognitive ability or intelligence in various ways. Elaborating on Spearman's notion of general cognitive ability (1927), Raven's advanced progressive matrices are found to measure fluid intelligence (Cattell, 1963), analytic intelligence (Carpenter et al., 1990), and intellectual efficiency if administered with a time limit (Hamel and Schmittmann, 2006).[6] As such, test scores on Raven's advanced progressive matrices can be interpreted as a proxy for cognitive ability. Indeed, the correlation between these test scores and students' grade point average (GPA) shows a significant and positive relationship in our sample.

Table 5.1 provides descriptive statistics of cognitive ability at the team level (panel A) and by field of study (panel B). Panel A indicates that the average number of figures correctly solved in the 20-minute timed version of Raven's advanced progressive matrices test is 18.60 out of 36 figures at maximum. We do not transform these test scores into an intelligence quotient, because this would require an additional assumption about the proper norm for first-year college students (Hamel and Schmittmann, 2006).[7]

[5]In contrast to the field experiments described in Chapters 3 and 4, the field experiment in this chapter did not administer a questionnaire halfway the entrepreneurship program.

[6]Spearman (1927) decomposed general cognitive ability (g) into an eductive and a reproductive component, where (i) eductive ability reflects "the ability to make meaning out of confusion, the ability to generate high-level, usually nonverbal, schemata which make it easy to handle complexity", and (ii) reproductive ability reflects "the ability to absorb, recall, and reproduce information that has been made explicit and communicated from one person to another" (Raven, 2000, p. 2). Fluid intelligence, analytic intelligence, intellectual efficiency and, hence, the 20-minute timed version of Raven's advanced progressive matrices test, mainly relate to eductive ability. Nevertheless, scores on cognitive tests such as Raven's advanced progressive matrices may differ across time, gender and culture (see Calvin et al., 2011; Irwing and Lynn, 2005; Nisbett et al., 2012; Raven, 2000; Rushton and Jensen, 2005).

[7]In line with Bors and Stokes (1998) and Raven et al. (1998) we also exclude students with a test score $\leqslant 6$ from the sample. Including this group of 10 students in total may incorrectly inflate teams' ability dispersion since a test score $\leqslant 6$, in more convential intelligence terms, roughly corresponds

Table 5.1. Descriptive statistics of cognitive ability

A: Cognitive ability	Mean	SD	Min	Max			
Average ability	18.60	2.53	14.00	23.22			
Ability dispersion (CV)	0.22	0.09	0.07	0.47			

B: Field of study	Stud.	Teams	Average ability			Ability dispersion		
			Mean	Min	Max	Mean	Min	Max
Business management	265	21	18.74	14.07	22.40	0.22	0.07	0.47
Management	45	4	18.67	16.27	22.22	0.23	0.18	0.33
Trade management Asia	108	10	18.92	14.78	23.22	0.20	0.10	0.35
Business languages	123	12	17.94	14.00	21.86	0.21	0.08	0.36
Financial management	32	2	19.33	18.83	19.83	0.27	0.21	0.32
Total	573	49	18.60	14.00	23.22	0.22	0.07	0.47

Note: Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test.

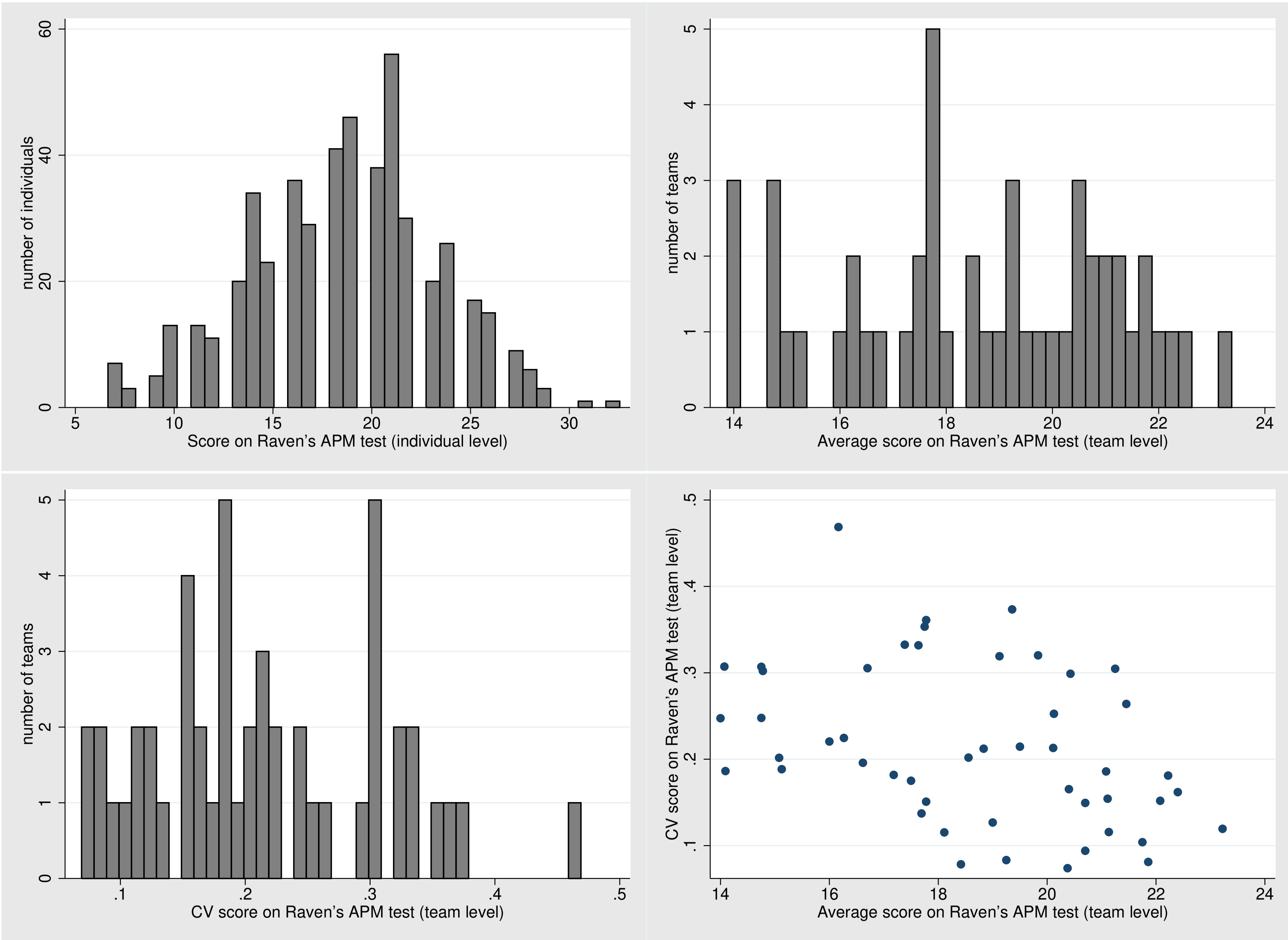
Moreover, our main interest is in the exogenous variation in cognitive ability rather than the exact level of test scores. Consistent with recent empirical studies involving professional sports (Franck and Nüesch, 2010; Papps et al., 2011), we use the coefficient of variation in test scores as a scale-invariant measure for ability dispersion in teams. Teams' coefficient of variation in test scores varies between 0.07 and 0.47 with a sample average of approximately 0.22. Panel B shows the numbers of students and teams by field of study. It also indicates that (the range of) average ability and ability dispersion of teams are similar across fields of study; possibly except for the field of financial management which accommodates only two teams.

Figure 5.2 shows frequency distributions of scores on Raven's advanced progressive matrices test at the individual and team level (average ability and ability dispersion). At the individual level test scores range from 7 to 32 figures correctly solved. The average ability of teams varies between test scores of 14 and 23, while as mentioned above teams' ability dispersion ranges from 0.07 to a coefficient of variation in test scores of 0.47.

One might worry that the effect of ability dispersion on team performance is biased since teams of low or high average ability, by construction, tend to have a lower ability dispersion (relative to teams of medium average ability). The scatter plot of teams' average ability and ability dispersion, however, does not reveal a systematic pattern that may confound a causal interpretation of the effect of ability dispersion on the business performance of teams (see Figure 5.2). Moreover, the results in Section 5.4

with the cognitive ability level of elementary school dropouts (which is highly unlikely for first-year college students). Students with a test score $\leqslant 6$ most likely just did not put in effort or choked while taking the test. T-tests acknowledge that those students are not significantly different from students with a test score > 6 in terms of age, gender, risk aversion and GPA.

Figure 5.2. Frequencies and scatter of scores on Raven's advanced progressive matrices test (individual and team level)



are similar if we include only 31 medium-ability teams with average test scores not more than one standard deviation away from the average test score in the sample, i.e., with average test scores in the range of 18.60 ± 2.53 (see Table 5.1).

5.3.2 Other variables

Table 5.2 provides descriptive statistics of individuals and teams. It shows that students are 20.9 years old on average and 43% of them are female. We also measure risk aversion (Dohmen et al., 2011) and students' scholastic achievements just before entering the college (indicated by 'grade point average'). At the team level, Table 5.2 indicates considerable variation in the number and incidence of dismissals across teams. In Section 5.4 we will examine to what extent dismissals vary between teams of different ability composition.[8]

Table 5.2. Descriptive statistics of individual and team characteristics

	Scale	Mean	SD	Min	Max
Individual level					
Age	years	20.93	2.09	17.89	32.86
Gender (dummy = 1 if female)	0/1	0.43	0.50	0.00	1.00
Risk aversion	1-11	7.45	2.00	1.00	11.00
Grade point average	1-10	6.38	0.28	5.90	6.98
Team level					
Size (at baseline)	persons	11.69	2.27	7.00	17.00
Dismissals (number)	number	0.35	0.60	0.00	2.00
Dismissals (incidence)	0/1	0.29	0.46	0.00	1.00

As in Chapter 4, business performance is operationalized by four measures: sales, profits, a binary indicator for positive profits and profits per share. Again, we include a binary indicator for positive profits to account for the fact that students tend to view as the bottom line result whether or not they are able to satisfy shareholders. The business performance of teams in this field experiment is roughly similar to the business performance of teams in the previous field experiments (see Chapters 3 and 4). Table 5.3 shows that average sales for all 49 teams are equal to 902 euros and that profits are 24 euros on average. More than half of the teams makes a profit (57%) and average profits per share amount to 0.62 euros. All three profit measures are significantly and positively related to sales.

If we split the sample into teams of low (mean<17), moderate (17≥mean≤21) and high (mean>21) average ability, descriptive statistics suggest that teams of moderate average ability perform slightly better on the different business outcomes. However,

[8] The number and incidence of dismissals were unrelated to the gender and ethnic diversity of teams in the field experiments of Chapters 3 and 4.

Table 5.3. Descriptive statistics of business outcomes

			Average ability			Ability dispersion		
		All	Low mean<17	Moderate 17≥mean≤21	High mean>21	Low CV<0.15	Moderate 0.15≥CV≤0.30	High CV>0.30
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sales (euros)	Mean	902	939	944	762	714	1140	636
	SD	974	909	1111	756	537	1267	478
	Min	0	43	0	42	210	42	0
	Max	5049	3505	5049	2553	2019	5049	1576
Profits (euros)	Mean	24	-18	91	-80	-94	115	-37
	SD	422	347	492	323	391	491	273
	Min	-993	-637	-993	-592	-993	-637	-596
	Max	1619	413	1619	353	508	1619	376
Positive profits (0/1)	Mean	0.57	0.62	0.60	0.45	0.42	0.63	0.62
	SD	0.50	0.51	0.50	0.52	0.51	0.49	0.51
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Profits per share (euros)	Mean	0.62	0.37	1.43	-0.91	-0.85	1.94	-0.46
	SD	7.01	5.71	7.88	6.58	5.31	8.42	5.27
	Min	-13.76	-9.65	-10.13	-13.76	-10.13	-13.76	-8.25
	Max	25.70	9.89	25.70	8.02	8.02	25.70	9.89
N		49	13	25	11	12	24	13

Note: Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test.

note that these descriptives are very sensitive to the exact location of particularly the second cutoff point. A cutoff at a mean test score of 20 (instead of 21), for example, would already imply that teams of high average ability achieve better results. If we split up the sample in teams of low (CV<0.15), moderate (0.15≤CV≤0.30) and high (CV>0.30) ability dispersion, teams of moderate ability dispersion tend to have higher sales, profits and profits per share than teams in the other two categories, on average. This ranking is rather insensitive to the precise location of the cutoff points. Section 5.4 examines the effect of the average level and dispersion of ability more formally.

5.3.3 Randomization

To assess whether the assignment of students to teams was truly random (conditional on their cognitive ability), we regress background characteristics of low ability and high ability students on the average test score in their team, the team's coefficient of variation in test scores and its square. This is consistent with the team level specifications of the main results in Section 5.4.

Table 5.4. Randomization checks at the individual and team level

	CV ability		CV ability^2		Average ability	
A1: Low ability students						
Age	-4.825	(5.894)	16.482	(10.468)	-0.029	(0.068)
Gender	1.591	(1.293)	-4.452*	(2.394)	0.029*	(0.016)
Risk aversion	-3.301	(6.071)	13.287	(12.401)	-0.048	(0.070)
Grade point average	0.227	(0.807)	-1.311	(1.462)	-0.003	(0.010)
A2: High ability students						
Age	7.163	(5.515)	-10.984	(10.448)	-0.057	(0.049)
Gender	-1.682	(1.992)	3.388	(4.653)	-0.020	(0.017)
Risk aversion	-1.639	(6.008)	5.894	(11.588)	0.009	(0.070)
Grade point average	0.755	(0.499)	-0.874	(0.991)	0.005	(0.006)
B: Team level (average)						
Age	-0.048	(5.651)	5.631	(11.485)	-0.009	(0.047)
Gender	-0.537	(2.333)	0.517	(5.366)	-0.009	(0.012)
Risk aversion	-3.195	(6.361)	10.764	(12.887)	-0.009	(0.048)
Grade point average	0.371	(0.593)	-0.789	(1.280)	0.020***	(0.005)
Team size	-1.531	(18.318)	8.077	(39.221)	-0.057	(0.135)

Note: Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test. In panels A1 and A2 each coefficient comes from a regression at the individual level of the row variable on the column variables, separately for students of low (test score ≤ 18.60) and high (test score > 18.60) cognitive ability (robust standard errors in parentheses). In panel B each coefficient comes from a regression at the team level of the row variable on the column variables (bootstrapped standard errors in parentheses; 1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

Panel A1 of Table 5.4 shows that background characteristics of low ability students do not systematically vary across teams of different ability composition. Hence, low ability students in teams of low ability dispersion are not significantly different from low ability students in teams of high ability dispersion. The same holds for background characteristics of high ability students (see panel A2). Low ability and high ability students assigned to teams of distinct ability composition are also not more or less likely to follow a specific field of study (not tabulated). In a similar fashion, panel B of Table 5.4 examines at the team level whether (average) background characteristics of students correlate with the ability composition of teams. Again, there are no systematic differences between teams of different ability composition.[9]

5.4 Results

5.4.1 Main findings

The key prediction of our proposed model in Section 5.2 is that team performance exhibits an inverse u-shaped relationship with ability dispersion. Another prediction of the model is that teams of moderate ability dispersion experience fewer dismissals due to fewer shirking members in those teams. This subsection presents the empirical findings in the order of the propositions of Section 5.2.

Table 5.5 reports regression results for the effect of ability dispersion on business performance as measured in terms of sales, profits, a binary indicator for positive profits and profits per share (see Proposition 1). In panel A these performance measures are regressed on teams' average test score, their coefficient of variation in test scores and its square (panel B provides results from spline functions). The linear effect of ability dispersion on business performance turns out insignificant in all specifications (not tabulated). Besides standard OLS regression, we employ median and robust (M-estimation) regression to assess whether the results are sensitive to outliers.

Column (1) shows that, given teams' average ability, sales first increase with ability dispersion up to a coefficient of variation in test scores of approximately 0.25 (the sample average is 0.22) and then decrease with ability dispersion. However, columns (2) and (3) indicate that this effect of ability dispersion on sales tends to be inflated by outliers: the point estimates are insignificant when using median and robust (M-estimation) regression. Columns (4) through (6) consistently show an inverse u-shaped pattern for the relationship between ability dispersion and profits. Again, performance

[9]Since the randomization checks in this subsection fail to find any pretreatment differences (that may contaminate the design of our field experiment), the analyses in the next section do not include control variables (adding superfluous controls would only reduce the degrees of freedom).

Table 5.5. Ability dispersion and team performance

	Sales			Profits			Pos. profits	Profits per share		
	OLS	Median	Robust	OLS	Median	Robust	OLS	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: Polynomial										
CV ability	14772.5*	4439.5	4422.5	8243.7**	7378.1**	6052.4*	8.952***	106.100*	92.478*	89.344
	(8083.3)	(5549.2)	(5217.3)	(3543.9)	(3200.0)	(3233.6)	(3.163)	(58.466)	(52.612)	(57.936)
CV ability^2	-30305.8*	-9485.4	-9118.5	-16025.3**	-14878.5**	-12441.4**	-16.677**	-207.870*	-196.543*	-178.400
	(17055.2)	(10975.2)	(11357.6)	(6795.2)	(6385.5)	(6276.4)	(6.580)	(112.727)	(104.931)	(115.048)
Avg. ability	25.529	3.449	-2.862	19.528	18.594	4.261	0.004	0.252	0.105	0.115
	(62.982)	(50.458)	(44.774)	(25.975)	(26.966)	(25.490)	(0.032)	(0.461)	(0.475)	(0.467)
Maximum	*0.24*	*0.23*	*0.24*	*0.26*	*0.25*	*0.24*	*0.27*	*0.26*	*0.24*	*0.25*
R^2	*0.09*	*0.04*	*0.03*	*0.15*	*0.08*	*0.11*	*0.14*	*0.09*	*0.08*	*0.07*
B: Spline (0.25)										
1st segment	7377.2*	2629.7	1976.8	3817.7**	3845.9**	2695.7	3.892***	51.036*	46.191	50.538
	(3788.4)	(2486.5)	(2310.0)	(1895.2)	(1583.8)	(1806.3)	(1.508)	(30.705)	(28.296)	(30.742)
2nd segment	-8317.5**	-2818.4	-2247.2	-3352.9**	-4022.5**	-2879.5*	-2.576	-47.116*	-59.334**	-46.582*
	(4060.6)	(3323.5)	(3133.5)	(1694.5)	(1560.1)	(1724.4)	(1.796)	(28.036)	(25.836)	(28.026)
Avg. ability	42.106	29.512	0.560	25.513	36.167	12.321	0.007	0.342	0.209	0.349
	(67.387)	(51.983)	(45.011)	(27.893)	(29.403)	(27.956)	(0.034)	(0.515)	(0.516)	(0.543)
R^2	*0.14*	*0.05*	*0.04*	*0.17*	*0.09*	*0.13*	*0.13*	*0.11*	*0.10*	*0.12*

Note: Based on information from 49 teams. Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test. OLS, Median and Robust refer to the estimation method. Median and robust specifications for positive profits are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

is maximized at a coefficient of variation in test scores of about 0.25. The same holds for the probability of profits being positive in column (7), although the degree of ability dispersion where performance peaks marginally increases to a coefficient of variation in test scores of 0.27. The coefficients in columns (8) through (10) corroborate these findings: the effect of ability dispersion on profits per share is described by an inverse u-shape with the optimum at a coefficient of variation in test scores of roughly 0.25. The results from the quadratic specifications in columns (4) through (10) are robust to outliers. Similar results are obtained when we exclude teams' average ability or include higher-order terms for the average ability of teams (not tabulated).

In panel B we estimate spline functions to address asymmetric effects of ability dispersion below (1st segment) and above (2nd segment) a coefficient of variation in test scores of 0.25. The cutoff in our spline functions is obtained by averaging the coefficients of variation in cognitive ability that maximize team performance (according to the quadratic specifications). Results from these spline functions indicate that business performance tends to increase with ability dispersion below a coefficient of variation in test scores of 0.25. If the coefficient of variation in test scores is at least equal to 0.25 all coefficients for the impact of ability dispersion are negative and (with three exceptions) significant. The point estimates in column (5) of panel B imply that raising the coefficient of variation in test scores from 0.20 to 0.25 increases profits by about 200 euros (approximately half of a standard deviation), while profits decrease by roughly the same amount if the coefficient of variation in test scores is further raised from 0.25 to 0.30. In sum, the results of panel B closely resemble the inverse u-shaped pattern from the quadratic specifications of panel A, although the number of teams may slightly limit the precision of its estimates.

Table 5.6 tests Proposition 2 by estimating the relationship between ability dispersion, dismissals and business performance. The number and incidence of dismissals reflect respectively the number of dismissals per team and whether or not a team has experienced at least one dismissal (dummy = 1 if any). Panel A shows for both number and incidence that teams of moderate ability dispersion are characterized by fewer dismissals and, we infer, less free-riding.[10] Dismissals do not reflect a process whereby teams simply get rid of low ability or high ability students (rather than shirkers) since the relationship between cognitive ability and probability of dismissal at the individual level is insignificant (not tabulated). Moreover, the number and incidence of dismissals are minimized at a coefficient of variation in test scores of approximately 0.24. Note that this minimum almost exactly corresponds with the coefficient of variation in test scores that maximizes business performance (about 0.25). Panel B indicates that fewer

[10]Interviews with program coordinators acknowledge that the main cause of dismissals is shirking of team members.

Table 5.6. Ability dispersion, dismissals and team performance

A: Dismissals		Number		Incidence
	OLS	Median	Robust	OLS
CV ability	-14.140***	-18.227**	-13.057**	-8.416***
	(4.154)	(8.399)	(5.523)	(3.260)
CV ability^2	29.525***	38.475**	28.681***	17.191***
	(8.136)	(17.091)	(10.264)	(6.489)
Avg. ability	0.018	-0.001	-0.007	0.012
	(0.027)	(0.020)	(0.028)	(0.024)
Minimum	*0.24*	*0.24*	*0.23*	*0.24*
R^2	*0.37*	*0.03*	*0.36*	*0.25*

B: Performance		Sales			Profits		Pos. profits		Profits per share	
	OLS	Median	Robust	OLS	Median	Robust	OLS	OLS	Median	Robust
Dismissals	-572.6***	-283.5	-278.9*	-193.0**	-176.0	-185.2*	-0.209	-3.270*	-4.392**	-3.120*
(number)	(189.1)	(202.5)	(164.4)	(90.5)	(127.4)	(98.8)	(0.129)	(1.775)	(1.988)	(1.874)
R^2	*0.13*	*0.05*	*0.10*	*0.07*	*0.05*	*0.10*	*0.07*	*0.07*	*0.06*	*0.08*
Dismissals	-774.2***	-355.2	-430.0**	-244.3**	-164.0	-223.4*	-0.259	-4.653**	-4.057	-4.038*
(incidence)	(269.1)	(285.5)	(208.7)	(123.7)	(166.8)	(127.8)	(0.166)	(2.033)	(2.768)	(2.137)
R^2	*0.14*	*0.06*	*0.14*	*0.06*	*0.03*	*0.09*	*0.06*	*0.08*	*0.04*	*0.08*

Note: Based on information from 49 teams. Average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test. All specifications control for team size. OLS, Median and Robust refer to the estimation method. Median and robust specifications for incidence of dismissals and positive profits are excluded since these variables are dichotomous. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

dismissals are also positively related to business performance, separately for number and incidence. However, we lack exogenous variation to identify a causal impact of dismissals on business performance, since dismissals are obviously endogenous.

To test whether knowledge breadth increases with cognitive ability, we regress the different measures of business performance on a third-order term for ability dispersion (see Proposition 3). Its coefficients turn out insignificant in all specifications (not tabulated), which suggests that knowledge breadth is invariant to cognitive ability in our setting. This is a novel finding, which is corroborated by the (non-tabulated) insignificant point estimates for the average ability of team members. Similar insignificant results are obtained if we regress business performance on average ability and a third-order term for ability dispersion simultaneously (not tabulated).

Consistent with the predictions of our model, the results presented in this subsection show that team performance exhibits an inverse u-shaped relationship with ability dispersion. With average cognitive ability held constant, team performance is maximized at a coefficient of variation in cognitive ability of approximately 0.25 (the sample average is 0.22). Teams of moderate ability dispersion also experience fewer dismissals during the program (i.e., lower degree of shirking members), although this does not chiefly explain why those teams perform better. In contrast to empirical studies which have analyzed settings involving laboratory experiments (Woolley et al., 2010), unskilled work tasks (e.g., Hamilton et al., 2003), or competitive sports (Kahn, 2000), we do not find evidence that average cognitive ability of team members significantly determines performance in a setting that requires coordination on a broad array of complex decision-making tasks entailing the sustained application of their cognitive abilities. Hence, our results suggest that (moderate) dispersion of cognitive ability trumps average cognitive ability in teams comprised of individuals of relatively high cognitive ability that have to complete a complicated business project (see Hong and Page, 2001, 2004). We do not find support for the notion that the lowest-ability member (i.e., the weakest link) or the highest-ability member (i.e., the superstar) in a team significantly affect team performance (Kremer, 1993; Prat, 2002). Results turn out being insignificant too when we relate teams' top 3 or bottom 3 students in terms of cognitive ability to the performance of teams.

5.4.2 Robustness

Robustness checks in this subsection are conducted by testing other measures of ability dispersion such as teams' standard deviation in ability (Hansen et al., 2006), teams' ratio of the maximum to the minimum ability (Hamilton et al., 2003), and spline functions with three segments of ability dispersion (cutoff levels at a coefficient of variation in test scores of 0.15 and 0.30).

Panels A and B of Table A1 in the appendix also reveal an inverse u-shaped effect of teams' standard deviation in ability and teams' ratio of the maximum to the minimum ability on their performance, although significance levels vary across both measures of ability dispersion.[11] The degree of ability dispersion that maximizes team performance is again slightly above the sample average.

Panel C indicates a similar inverse u-shaped pattern for spline functions with teams of low, moderate and high ability dispersion (based on the coefficient of variation in test scores). The point estimates for teams of low ability dispersion (CV$<$0.15) are positive and relatively large compared to the coefficients for teams of moderate ability dispersion (0.15$\geq$CV$\leq$0.30). For teams of high ability dispersion (CV$>$0.30) the point estimates are negative and also relatively large in relation to those for teams of moderate ability dispersion. The number of teams, however, limits the precision of these estimates. In sum, none of the robustness checks are at odds with the results previously obtained.

5.5 Discussion and conclusion

We have studied the effect on organizational performance of a team's composition in terms of cognitive ability, a major determinant of economic behavior and outcomes (e.g., Hanushek and Woessmann, 2008). We propose a model in which greater ability dispersion generates greater knowledge for a team, but also increases the costs of monitoring necessitated by moral hazard.

In line with the predictions of our model, team performance exhibits an inverse u-shaped relationship with ability dispersion. Controlling for the average cognitive ability of teams, performance is maximized at a coefficient of variation in cognitive ability of about 0.25 (the sample average is 0.22). Teams of moderate ability dispersion also experience fewer dismissals due to a lower degree of free-riding members, although this does not chiefly explain why those teams perform better.

There is ample opportunity to extend our model, for example, by allowing for richer interactions between team members. That might enable researchers to study other interesting questions such as preferences, beliefs and (re)negotiation in teams. Other limitations relate to the experimental set-up of our study. We exploit exogenous variation in cognitive ability among students who probably lack serious work experience, which may limit the external validity of our findings.

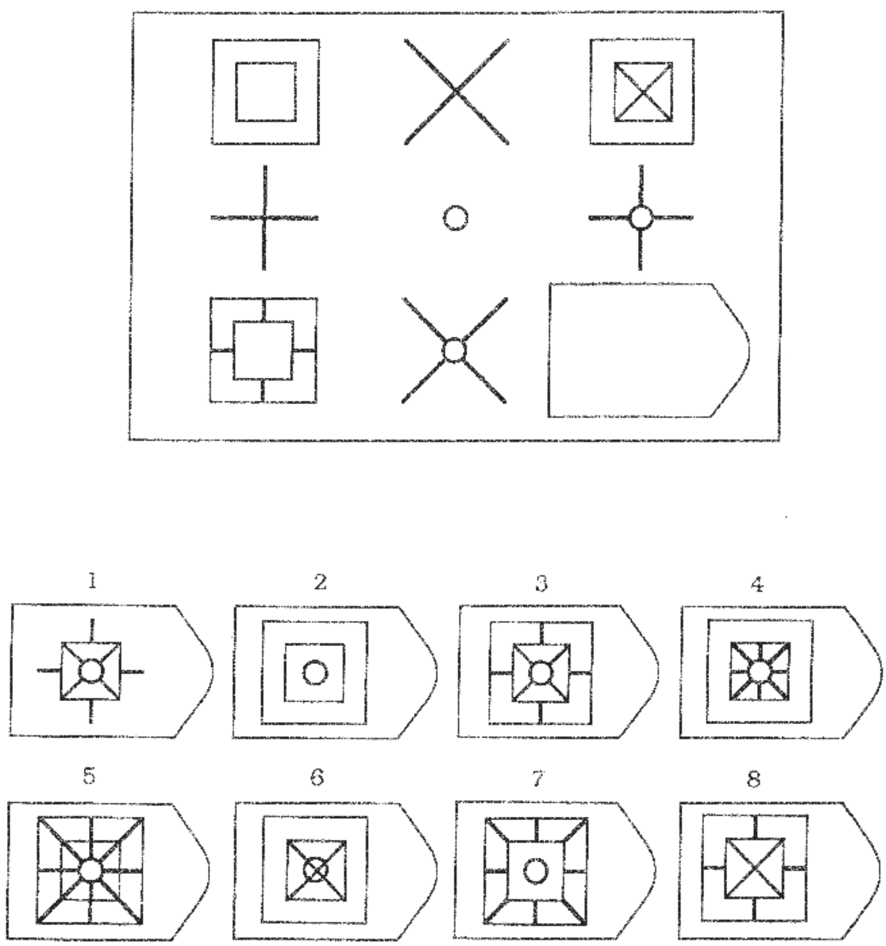
Nevertheless, students in our experiment execute a substantial business project in which tasks are diverse and involve complex decision-making entailing the sustained

[11] The ratio of the maximum to the minimum ability in the team is possibly more sensitive to outliers since this measure of ability dispersion could already be considerably inflated by only one team member of (very) low or high cognitive ability.

application of their cognitive abilities. Moreover, students face strong incentives that align their interests with those of the team. All in all, we therefore have grounds to believe that our field experiment is informative about the impact of ability dispersion on the performance of business (management) teams. A next step for future research would be to replicate experiments like this for teams with various kinds of tasks that differ in terms of their degree of diversity and complexity.

Appendix Figure A1. Example of a figure from Raven's advanced progressive matrices test

19



Appendix Table A1. Other measures of ability dispersion and team performance

	Sales			Profits			Pos. profits	Profits per share		
	OLS	Median	Robust	OLS	Median	Robust	OLS	OLS	Median	Robust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: Standard deviation										
SD ability	955.5**	370.4	239.7	548.5***	538.0**	412.6**	0.641***	7.529**	6.324*	6.010**
	(446.4)	(286.2)	(272.0)	(202.6)	(206.4)	(169.9)	(0.195)	(3.131)	(3.650)	(2.765)
SD ability^2	-107.6**	-43.7	-27.7	-59.4***	-65.1**	-46.9**	-0.068***	-0.827**	-0.801*	-0.679**
	(48.4)	(33.4)	(32.3)	(22.0)	(24.6)	(19.2)	(0.024)	(0.339)	(0.424)	(0.314)
Avg. ability	18.841	30.451	-6.546	11.389	-2.816	-0.248	-0.008	0.163	-0.083	0.013
	(57.584)	(44.307)	(36.635)	(20.706)	(23.714)	(18.422)	(0.027)	(0.386)	(0.462)	(0.364)
Max./Avg.	*4.44/3.91*	*4.24/3.91*	*4.33/3.91*	*4.62/3.91*	*4.13/3.91*	*4.40/3.91*	*4.71/3.91*	*4.55/3.91*	*3.95/3.91*	*4.43/3.91*
R^2	*0.08*	*0.03*	*0.02*	*0.15*	*0.06*	*0.12*	*0.16*	*0.10*	*0.07*	*0.08*
B: Ratio maximum/minimum										
Ratio ability	2208.6**	1037.6	709.8	843.1	386.5	512.0	0.849	10.392	7.459	7.608
	(1043.0)	(787.2)	(788.1)	(550.5)	(571.2)	(606.1)	(0.683)	(9.255)	(10.154)	(10.879)
Ratio ability^2	-403.5*	-181.8	-131.9	-151.5	-58.3	-91.0	-0.140	-1.851	-1.127	-1.342
	(212.2)	(164.0)	(167.3)	(114.9)	(124.1)	(130.6)	(0.146)	(1.908)	(2.214)	(2.352)
Avg. ability	43.312	40.983	7.209	18.076	13.565	3.956	0.003	0.242	0.389	0.089
	(56.444)	(41.184)	(41.247)	(23.082)	(26.870)	(22.130)	(0.033)	(0.438)	(0.530)	(0.436)
Max./Avg.	*2.74/2.17*	*2.85/2.17*	*2.69/2.17*	*2.78/2.17*	*3.31/2.17*	*2.81/2.17*	*3.03/2.17*	*2.81/2.17*	*3.31/2.17*	*2.83/2.17*
R^2	*0.10*	*0.05*	*0.04*	*0.08*	*0.03*	*0.04*	*0.08*	*0.04*	*0.03*	*0.03*
C: Spline (0.15 and 0.30)										
1st segment	9022.6*	2193.0	186.9	5166.3*	5085.9	3918.4	7.297**	55.414	60.502	51.447
(N=12)	(4649.3)	(4643.2)	(4629.7)	(2702.6)	(3112.3)	(2754.7)	(2.960)	(40.254)	(53.613)	(42.579)
2nd segment	1045.1	408.3	1805.1	1120.6	481.3	441.3	0.759	17.213	1.709	10.826
(N=24)	(3232.4)	(2600.9)	(2350.0)	(1493.7)	(1521.2)	(1687.2)	(1.814)	(27.227)	(32.648)	(31.896)
3rd segment	-8014.1	-2887.7	-4484.7	-4162.1	-4548.3	-3429.2	-2.799	-57.507	-61.100	-48.734
(N=13)	(9125.7)	(7342.2)	(6849.5)	(3329.1)	(4084.8)	(4576.6)	(4.542)	(65.005)	(71.815)	(109.826)
Avg. ability	21.255	8.269	-3.939	17.804	10.606	3.567	0.001	0.227	0.130	0.085
	(65.578)	(49.193)	(42.950)	(26.909)	(29.171)	(28.291)	(0.033)	(0.477)	(0.550)	(0.511)
R^2	*0.08*	*0.04*	*0.06*	*0.14*	*0.07*	*0.10*	*0.14*	*0.08*	*0.08*	*0.07*

Note: Based on information from 49 teams. In panel A average and SD of ability reflect at the team level respectively average score and standard deviation in scores on Raven's advanced progressive matrices test. In panel B average and ratio of ability reflect at the team level respectively average score and ratio of the maximum to the minimum score on Raven's advanced progressive matrices test. In panel C average and CV of ability reflect at the team level respectively average score and coefficient of variation in scores on Raven's advanced progressive matrices test (with cutoffs at a coefficient of variation in test scores of 0.15 and 0.30). OLS, Median and Robust refer to the estimation method. Median and robust specifications for positive profits are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1000 replications). ***/**/* denotes significance at the 1%/5%/10%-level.

Chapter 6

Concluding remarks

This dissertation has studied the impact of diversity on team performance using a series of field experiments in which teams start up and manage real companies under identical circumstances. Exogenous variation in - otherwise random - team composition is imposed by assigning individuals to teams based on their gender, ethnicity or cognitive ability. The setting of these field experiments closely resembles that of (business) management practices in the longer run where tasks are diverse and involve complex decision-making. Evidence from this kind of experiments potentially contributes to the effective composition of teams in organizations. The results of this dissertation demonstrate that diversity in gender, ethnicity and cognitive ability has substantial and non-monotonic effects on the performance of teams. Various underlying mechanisms are explored to explain why teams of distinct degrees of diversity perform differently.

The first field experiment on the impact of gender diversity indicates that teams with an equal gender mix perform better than male-dominated teams. In addition, the results suggest that gender diverse teams perform no worse or even slightly better than teams with a majority of females. However, the sample did not include enough female-dominated teams to conclude firmly that these results are symmetric. None of the underlying mechanisms forthcoming from the literature, and insofar as these mechanisms could be tested, are found to explain this positive effect of gender diversity in teams on their performance.

The effect of ethnic diversity on team performance was investigated in the second field experiment. This experiment shows that the relationship between the performance of teams and ethnic diversity is flat or tends to decline down to a certain threshold level and starts increasing beyond this threshold level (at which at least about half of the team is ethnically diverse). Hence, the net effect of ethnic diversity in teams on their performance only turns out positive if ethnic diversity is sufficiently substantial. In line with theoretical predictions, ethnically diverse teams tend to be characterized by a more diverse pool of relevant knowledge. Moreover, these teams experience more (mutual) learning which is probably facilitated by their more diverse pool of relevant knowledge. In sum, these findings suggest that complementarities in knowledge and more effective mutual learning are the underlying mechanisms that possibly explain superior performance of ethnically diverse teams. Unfortunately, the causality of the complementarities/learning-performance relationship cannot be established based on the research design that is employed in this dissertation.

The third field experiment has examined the impact of diversity in cognitive ability. It demonstrates that team performance exhibits an inverse u-shaped pattern with diversity in cognitive ability (controlling for the average cognitive ability of teams). This finding is consistent with the predictions of the proposed model in which diversity in cognitive ability has both beneficial and detrimental effects. In the model, greater diversity in cognitive ability generates greater knowledge for a team and, hence, more valuable team output. However, greater diversity in cognitive ability simultaneously reduces the effectiveness of monitoring to detect and punish shirking by team members. Monitoring of each others effort becomes more difficult to judge the more distinct team members are in terms of their cognitive ability. At sufficiently high degrees of diversity in cognitive ability, the probability of being caught and dismissed for shirking declines so much that shirking reduces the performance of teams. Hence, teams of moderate diversity in cognitive ability are associated with maximal performance. Teams with a moderate degree of diversity in cognitive ability also experience fewer dismissals, although this does not chiefly explain why those teams achieve better results.

The results of this dissertation may have implications for the optimal formation of teams in organizations (such as business, start-up and top management teams). Many decisions in organizations are nowadays assigned to teams, not to individuals. A better understanding of the determinants of the effectiveness of teams has therefore become increasingly relevant. One of the most salient determinants of the effectiveness of teams is their diversity. In addition, recent developments have strengthened the relevance of diversity for organizations: (i) the share of women in higher education and the labor force has increased, (ii) local populations have become more ethnically diverse in a globalized world and the share of ethnic minorities in Western populations is rising sharply, and (iii) performance increasingly depends on human capital and the accumulation of knowledge in information-based economies. The results of this dissertation indicate that diversity can be employed as an organizational tool to promote efficiency or profitability. Yet, evidence from this dissertation also shows that the precise impact of diversity in teams on their performance is subtle. As such, this dissertation demonstrates that diversity is indeed a valuable asset for organizations and a promising field of research for academics.

It is important to emphasize the role of the setting in which teams execute their tasks. Teams in the field experiments of this dissertation operate in an entrepreneurial setting that requires coordination on a broad array of complex decision-making tasks. This does not necessarily imply that the results based on this setting generalize to other sets of tasks with possibly different degrees of broadness and complexity. A similar reasoning applies to another potential determinant of team performance, namely the size of teams. The effects of diversity reported in this dissertation may not hold for teams of a different size. In addition, other dimensions of diversity that remained unexplored here (such as diversity in age, experience or personality) could reveal different effects on the performance of teams. This dissertation has also focused purposely on the separate and unconfounded impact of various dimensions of diversity in teams on their performance. Hence, the possible interplay between different dimensions of diversity was neglected. Until today, however, studies analyzing the causal effect of diversity on team performance in the longer run have been scarce. This dissertation has exploited the novel opportunity to conduct field experiments with stratified randomization that closely resemble the functioning of real-world teams and, hence, potentially improve our understanding of the impact of diversity on the performance of teams in public and private organizations.

A next step for future research would be to replicate experiments like in this dissertation, preferably in real organizations and on a larger scale. Follow-up studies may also direct their focus to the effects of diversity in teams of a different size or in other settings that involve, for example, unskilled work tasks or competitive sports. Moreover, future research could examine the impact of different dimensions of diversity (rather than diversity in gender, ethnicity or cognitive ability) and the interplay between distinct dimensions of diversity. Finally, experimental replication studies may analyze the effects of diversity in teams on other outcomes such as organizational identity, employee motivation or customer satisfaction.

Bibliography

Adams, R. and Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, 94(2):291–309.

Adams, R. and Funk, P. (2012). Beyond the glass ceiling: Does gender matter? *Management Science*, 58(2):219–235.

Adams, R., Hermalin, B., and Weisbach, M. (2010). The role of boards of directors in corporate governance: A conceptual framework and survey. *Journal of Economic Literature*, 48(1):58–107.

Ahern, K. and Dittmar, A. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics*, 127(1):137–197.

Akerlof, G. and Kranton, R. (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, 19(1):9–32.

Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Allen, J., Van Breugel, G., Coenen, J., Fouarge, D., Meng, C., Ramaekers, G., Van Dijk, J., and Venhorst, V. (2009). Afgestudeerden van het HBO tijdens een crisis: Geen verloren generatie. Technical report, HBO-raad.

Amason, A., Shrader, R., and Tompson, G. (2006). Newness and novelty: Relating top management team composition to new venture performance. *Journal of Business Venturing*, 21(1):125–148.

Ancona, D. and Caldwell, D. (1992). Demography and design: Predictors of new product team performance. *Organization Science*, 3(3):321–341.

Angrist, J. and Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94(5):1613–1634.

Angrist, J. and Pischke, J. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Apesteguia, J., Azmat, G., and Iriberri, N. (2012). The impact of gender composition on team performance and decision-making: Evidence from the field. *Management Science*, 58(1):78–93.

Aslund, O., Edin, P., and Fredriksson, P. (2011). Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy. *American Economic Journal: Applied Economics*, 3(2):67–95.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., and Sugden, R. (2010). *Experimental Economics: Rethinking the Rules*. Princeton: Princeton University Press.

Baumol, W. and Strom, R. (2010). *"Useful Knowledge" of Entrepreneurship: Some Implications of the History*, chapter in: D. Landes, J. Mokyr, and W. Baumol (eds.), The Invention of Enterprise: Entrepreneurship from Ancient Mesopotamia to Modern Times, pages 528–541. Princeton: Princeton University Press.

Bell, S. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, 92(3):595–614.

Bernstein, R. and Edwards, T. (2008). An older and more diverse nation by midcentury. Technical report, US Census Bureau.

Blau, P. (1977). *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.

Boisjoly, J., Duncan, G., Kremer, M., Levy, D., and Eccles, J. (2006). Empathy or antipathy? The impact of diversity. *American Economic Review*, 96(5):1890–1905.

Bors, D. and Stokes, T. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, 58(3):382–398.

Calvin, C., Deary, I., Fenton, C., Roberts, B., Der, G., Leckenby, N., and Batty, G. (2011). Intelligence in youth and all-cause-mortality: Systematic review with meta-analysis. *International Journal of Epidemiology*, 40(3):626–644.

Card, D. and Rothstein, J. (2007). Racial segregation and the black-white test score gap. *Journal of Public Economics*, 91(11-12):2158–2184.

Carpenter, P., Just, M., and Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, 97(3):404–431.

Carter, D., D'Souza, F., Simkins, B., and Simpson, G. (2010). The gender and ethnic diversity of US boards and board committees and firm financial performance. *Corporate Governance: An International Review*, 18(5):396–414.

Catalyst (2010). Catalyst census of women board directors of the Fortune 1000. Technical report, Catalyst.

Cattell, R. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1):1–22.

Charness, G., Rigotti, L., and Rustichini, A. (2007). Individual behavior and group membership. *American Economic Review*, 97(4):1340–1352.

Chen, Y. and Li, S. (2009). Group identity and social preferences. *American Economic Review*, 99(1):431–457.

Chowdhury, S. (2005). Demographic diversity for building an effective entrepreneurial team: Is it important? *Journal of Business Venturing*, 20(6):727–746.

Cutler, D. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1):1–28.

De Jong, A., de Ruyter, K., and Wetzels, M. (2005). Antecedents and consequences of group potency: A study of self-managing service teams. *Management Science*, 51(11):1610–1625.

Desvaux, G., Devillard-Hoellinger, S., and Baumgarten, P. (2007). Women matter: Gender diversity, a corporate performance driver. Technical report, McKinsey & Company.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–60.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dufwenberg, M. and Muren, A. (2006). Gender composition in teams. *Journal of Economic Behavior and Organization*, 61(1):50–54.

Dugar, S. and Shahriar, Q. (2012). Group identity and the moral hazard problem: Experimental evidence. *Journal of Economics & Management Strategy*, 21(4):1061–1081.

Earley, P. and Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal*, 43(1):26–49.

Edwards, B., Day, E., Arthur Jr., W., and Bell, S. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology*, 91(3):727–736.

Ellison, S., Greenbaum, J., and Mullin, W. (2010). Diversity, social goods provision, and performance in the firm. CESifo Working Paper.

Ensley, M. and Hmieleski, K. (2005). A comparative study of new venture top management team composition, dynamics and performance between university-based and independent start-ups. *Research Policy*, 34(7):1091–1105.

Falk, A. and Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57.

Farrell, K. and Hersch, P. (2005). Additions to corporate boards: The effect of gender. *Journal of Corporate Finance*, 11(2):85–106.

Franck, E. and Nüesch, S. (2010). The effect of talent disparity on team productivity in soccer. *Journal of Economic Psychology*, 31(2):218–229.

Goldberg, L. (1990). An alternative description of personality: The big five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216–1229.

Gould, E. and Winter, E. (2009). Interactions between workers and the technology of production: Evidence from professional baseball. *Review of Economics and Statistics*, 91(1):188–200.

Grinblatt, M., Keloharju, M., and Linnainmaa, J. (2011). IQ and stock market participation. *Journal of Finance*, 66(6):2121–2164.

Guiso, L., Sapienza, P., and Zingales, L. (2009). Cultural biases in economic exchange? *Quarterly Journal of Economics*, 124(3):1095–1131.

Gul, F., Srinidhi, B., and Ng, A. (2011). Does board gender diversity improve the informativeness of stock prices? *Journal of Accounting and Economics*, 51(3):314–338.

Hamel, R. and Schmittmann, V. (2006). The 20-minute version as a predictor of the Raven advanced progressive matrices test. *Educational and Psychological Measurement*, 66(6):1039–1046.

Hamilton, B., Nickerson, J., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497.

Hamilton, B., Nickerson, J., and Owan, H. (2012). *Diversity and Productivity in Production Teams*, chapter in: A. Bryson (ed.), Advances in the Economic Analyses of Participatory & Labor-Managed Firms, volume 13, pages 99–138. Bingley: Emerald Group Publishing Limited.

Hansen, Z., Owan, H., and Pan, J. (2006). The impact of group diversity on performance and knowledge spillover: An experiment in a college classroom. NBER Working Paper.

Hanushek, E., Kain, J., and Rivkin, S. (2009). New evidence about brown v. board of education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383.

Hanushek, E. and Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3):607–668.

Harrison, G. and List, J. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Hartog, J. (2011). *Ethnic Minorities in the Netherlands*, chapter in: M. Kahanec and K. Zimmerman (eds.), Ethnic Diversity in European Labor Markets: Challenges and Solutions, pages 198–221. Cheltenham: Edward Elgar Publishing.

HBO-raad (2010). Feiten en cijfers: Afgestudeerden en uitvallers in het hoger beroepsonderwijs. Technical report, HBO-raad.

Hermalin, B. and Weisbach, M. (2003). Boards of directors as an endogenously determined institution: A survey of the economic literature. *Economic Policy Review*, 9(1):7–26.

Herring, C. (2009). Does diversity pay? Race, gender, and the business case for diversity. *American Sociological Review*, 74(2):208–224.

Hong, L. and Page, S. (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory*, 97(1):123–163.

Hong, L. and Page, S. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 101(46):16385–16389.

Horwitz, S. and Horwitz, I. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33(6):987–1015.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper.

Iranzo, S., Schivardi, F., and Tosetti, E. (2008). Skill dispersion and firm productivity: An analysis with employer-employee matched data. *Journal of Labor Economics*, 26(2):247–285.

Irwing, P. and Lynn, R. (2005). Sex differences in means and variability on the progressive matrices in university students: A meta-analysis. *British Journal of Psychology*, 96(4):505–524.

Ivanova-Stenzel, R. and Kuebler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, 32(5):797–808.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: Assessing the impacts of cultural and language diversity in the national hockey league. *Review of Economics and Statistics*, forthcoming.

Kahn, L. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives*, 14(3):75–94.

Kandel, E. and Lazear, E. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–817.

Karlan, D. and Valdivia, M. (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics and Statistics*, 93(2):510–527.

Kremer, M. (1993). The O-ring theory of economic development. *Quarterly Journal of Economics*, 108(3):551–575.

Laffont, J. and Martimort, D. (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton: Princeton University Press.

Langfred, C. (2004). Too much of a good thing? Negative effects of high trust and individual autonomy in self-managing teams. *Academy of Management Journal*, 47(3):385–399.

Lavy, V. and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33.

Lazear, E. (1999). Globalisation and the market for team-mates. *Economic Journal*, 109(454):15–40.

Lee, C. and Farh, J. (2004). Joint effects of group efficacy and gender diversity on group cohesion and performance. *Applied Psychology*, 53(1):136–154.

Lee, N. and Nathan, M. (2011). Does cultural diversity help innovation in cities: Evidence from London firms. SERC Discussion Paper.

Leonard, J., Levine, D., and Giuliano, L. (2010). Customer discrimination. *Review of Economics and Statistics*, 92(3):670–678.

LePine, J. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology*, 88(1):27–39.

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology*, 88(4):587–604.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–145.

Mateos, P. (2007). A review of name-based ethnicity classification methods and their potential in population studies. *Population, Space and Place*, 13(4):243–263.

Matsa, D. and Miller, A. (2010). A female style in corporate leadership? Evidence from quotas. IZA Discussion Paper.

Mills, C., Ablard, K., and Brody, L. (1993). The Raven's progressive matrices: Its usefulness for identifying gifted/talented students. *Roeper Review*, 15(3):183–186.

Minniti, M. and Bygrave, W. (2001). A dynamic model of entrepreneurial learning. *Entrepreneurship Theory & Practice*, 25(3):5–16.

Montalvo, J. and Reynal-Querol, M. (2005). Ethnic diversity and economic development. *Journal of Development Economics*, 76(2):293–323.

Morgan, J. and Vardy, F. (2009). Diversity in the workplace. *American Economic Review*, 99(1):472–485.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101.

Nisbett, R., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D., and Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, 67(2):1–30.

Nygaard, K. (2011). Forced board changes: Evidence from Norway. NHH Discussion paper.

Oliver, R. and Anderson, E. (1994). An empirical test of the consequences of behavior- and outcome-based sales control systems. *Journal of Marketing*, 58(4):53–67.

Oosterbeek, H. and Van Ewijk, R. (2010). Gender peer effects in university: Evidence from a randomized experiment. Tinbergen Institute Discussion Paper.

Oosterbeek, H., Van Praag, M., and IJsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, 54(3):442–454.

Ottaviano, G. and Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, 6(1):9–44.

Ozgen, C., Nijkamp, P., and Poot, J. (2011a). Immigration and innovation in European regions. IZA Discussion Paper.

Ozgen, C., Nijkamp, P., and Poot, J. (2011b). The impact of cultural diversity on innovation: Evidence from Dutch firm-level data. IZA Discussion Paper.

Page, S. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton: Princeton University Press.

Papps, K., Bryson, A., and Gomez, R. (2011). Heterogeneous worker ability and team-based production: Evidence from major league baseball, 1920–2009. *Labour Economics*, 18(3):310–319.

Parker, S. (2009). *The Economics of Entrepreneurship*. Cambridge: Cambridge University Press.

Parrotta, P., Pozzoli, D., and Pytlikova, M. (2010). Does labor diversity affect firm productivity? Norface Migration Discussion Paper.

Pearsall, M., Ellis, A., and Evans, J. (2008). Unlocking the effects of gender faultlines on team creativity: Is activation the key? *Journal of Applied Psychology*, 93(1):225–234.

Pelled, L. (1996). Demographic diversity, conflict, and work group outcomes: An intervening process theory. *Organization Science*, 7(6):615–631.

Pelled, L., Eisenhardt, K., and Xin, K. (1999). Exploring the black box: An analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly*, 44(1):1–3.

Prat, A. (2002). Should a team be homogeneous? *European Economic Review*, 46(7):1187–1207.

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1):1–48.

Raven, J., Raven, J., and Court, J. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 1 General overview*. Oxford: Psychologists Press.

Raven, J., Raven, J., and Court, J. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 4 Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press.

Richard, O., Barnett, T., Dwyer, S., and Chadwick, K. (2004). Cultural diversity in management, firm performance, and the moderating role of entrepreneurial orientation dimensions. *Academy of Management Journal*, 47(2):255–266.

Rosen, S. (1981). The economics of superstars. *American Economic Review*, 71(5):845–858.

Rushton, J. and Jensen, A. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy and Law*, 11(2):235–294.

Shane, S., Nicolaou, N., Cherkas, L., and Spector, D. (2010). Genetics, the big five, and the tendency to be self-employed. *Journal of Applied Psychology*, 95(6):1154–1162.

Shore, L., Chung-Herrera, B., Dean, M., Ehrhart, K., Jung, D., Randel, A., and Singh, G. (2009). Diversity in organizations: Where are we now and where are we going? *Human Resource Management Review*, 19(2):117–133.

Spearman, C. (1927). *The Abilities of Man*. New York: Macmillan.

Stewart, G. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, 32(1):29–55.

Wegge, J., Roth, C., Neubach, B., Schmidt, K., and Kanfer, R. (2008). Age and gender diversity as determinants of performance and health in a public organization: The role of task complexity and group size. *Journal of Applied Psychology*, 93(6):1301–1313.

Woods, D. (2010). Female representation on UK boards has risen to 13.6 percent. HR Magazine, United Kingdom.

Woolley, A., Chabris, C., Pentland, A., Hashmi, N., and Malone, T. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686–688.

Zhao, H. and Seibert, S. (2006). The big five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology*, 91(2):259–271.

Samenvatting (Summary in Dutch)

Introductie

In de afgelopen decennia zijn teams een steeds grotere rol gaan spelen binnen organisaties. Activiteiten en beslissingen die vroeger vaak aan individuen werden toebedeeld worden steeds vaker neergelegd bij teams (Hamilton et al., 2003; Woolley et al., 2010). Denk bijvoorbeeld aan raden van bestuur. Hier wordt tegenwoordig veel meer met een collectief besturingsmodel gewerkt dan onder leiding van de voorzitter. Maar ook binnen de rechterlijke macht worden vonnissen tegenwoordig vaak door teams van rechters gewezen. Een ander voorbeeld is de wetenschap waarin steeds meer artikelen met teams van onderzoekers worden geschreven. Ook worden de meeste startende ondernemingen tegenwoordig opgezet door teams in plaats van eenlingen (Parker, 2009).

Deze ontwikkelingen hebben te maken met de toenemende rol van in hoog tempo verouderende kennis. Daardoor zijn tijd en specialistische kennis bij het nemen van beslissingen belangrijker geworden. Teams kunnen ten opzichte van individuen in kortere tijd op basis van een breder scala aan actuele kennis beslissen en activiteiten ontplooien. Een van de karakteristieken van teams is hun samenstelling. Meer inzicht in de effectieve samenstelling van teams kan hierdoor een waardevol instrument voor organisaties zijn. Het is ook een boeiend onderzoeksterrein voor wetenschappers geworden (Hamilton et al., 2012). Een van de meest karakteristieke kenmerken van teams in organisaties is hun diversiteit (Page, 2007).

Diversiteit binnen organisaties wordt vaak geassocieerd met voordelen zoals een meer diverse verzameling van talenten. Maar ook wordt diversiteit verbonden met het imago en de identiteit van organisaties. Bovendien worden aan diversiteit kwaliteiten toegedicht die zouden leiden tot betere relaties met belanghebbenden, een meer gemotiveerd werknemersbestand en een grotere klanttevredenheid (bijv. Desvaux et al., 2007). Adviesbureaus zoals The Boston Consulting Group en McKinsey & Company pleiten voor diversiteit als middel om de effectiviteit of winstgevendheid van organisaties te vergroten. Diversiteit is een relevant onderwerp voor managers, beleidsmakers en belangengroepen geworden.

Andere argumenten die vaak worden genoemd in de discussie rondom diversiteit zijn meer maatschappelijk geöriënteerd en hebben niet direct te maken met de prestaties van organisaties. Denk aan het betrekken van (minderheids)groepen in het maatschappelijk verkeer, het nastreven van een representatieve afspiegeling van de samenleving in organisaties en de vermindering van stereotypen (door omgang met andersdenkenden).

In het wetenschappelijk debat wordt diversiteit vaak in verband gebracht met baten zoals complementariteiten en de mogelijkheid om van anderen te leren. Een meer divers team brengt een bredere set aan vaardigheden en kennis mee. Daardoor vullen teamleden elkaar eerder aan en kunnen ze meer van elkaar leren. Dit kan leiden tot meer creativiteit en een betere besluitvorming. Maar diversiteit kan ook zorgen voor hogere kosten doordat communicatie en coördinatie moeilijker worden. Naarmate individuen meer verschillen in hun handel-, denk- of communicatiewijze is er minder snel begrip en kan een slechtere sfeer ontstaan die mogelijk zelfs leidt tot conflicten (Lazear, 1999).

Het is onduidelijk of de kosten dan wel de baten van diversiteit groter zijn in verschillende contexten. Het netto effect van diversiteit op de prestaties van teams blijft dan ook een vraag die alleen door empirisch onderzoek kan worden beantwoord. Het verband tussen de prestaties van teams en hun diversiteit is uitgebreid onderzocht: (i) in verschillende vakgebieden zoals economie en financiering, ondernemerschap, bedrijfs- en organisatiekunde, psychologie en sociologie; (ii) voor verschillende dimensies van diversiteit zoals (cognitieve) vaardigheden, leeftijd, etniciteit, ervaring, geslacht, normen en persoonlijkheid; (iii) op verschillende niveaus zoals complete organisaties, teams op de werkvloer, raden van bestuur en teams van ondernemers (zie Shore et al., 2009).

De grote meerderheid van studies in dit onderzoeksveld meet het *verband* tussen diversiteit en de prestaties van teams in plaats van het *effect* van diversiteit in teams op hun prestaties. De gemeten verbanden in onderzoek aan de hand van teams uit de praktijk geven waarschijnlijk een vertekend beeld. De samenstelling van teams in organisaties wordt immers vaak bepaald door verwachtingen van managers over de productiviteit of door zelfselectie van werknemers in en uit teams (zie het voorbeeld hieronder). De daadwerkelijke effecten van diversiteit op de prestaties van teams blijven op deze manier onbekend (Adams et al., 2010; Hermalin en Weisbach, 2003).

De meeste studies meten daarnaast verschillende dimensies van diversiteit tegelijkertijd. Dit maakt een analyse van iedere afzonderlijke dimensie van diversiteit onmogelijk. Het specifieke effect van een bepaalde dimensie van diversiteit in teams op hun prestaties blijft hiermee onduidelijk. Wat is bijvoorbeeld het afzonderlijke effect van diversiteit in geslacht op de prestaties van teams zonder gelijktijdig de effecten van diversiteit in leeftijd of etniciteit te meten?

Het doel van dit proefschrift is het een voor een ontrafelen van de effecten van verschillende dimensies van diversiteit in teams op hun prestaties. Hiervoor zijn drie veldexperimenten uitgevoerd die de afzonderlijke effecten van diversiteit in geslacht, etniciteit en cognitieve vaardigheden meten. De resultaten van dit type onderzoek zijn waardevol voor een beter inzicht in de effectieve samenstelling van teams in organisaties.

Het belang van het meten van het *effect* van diversiteit op de prestaties van teams kan worden verduidelijkt aan de hand van de huidige discussie rondom een meer gelijke verdeling tussen mannen en vrouwen in raden van bestuur (en raden van commissarissen). Argumenten die voor meer diversiteit pleiten zijn vaak gebaseerd op studies die een positief verband vinden tussen het aandeel vrouwen in raden van bestuur en de prestaties van het bedrijf. Maar dit positieve verband hoeft niet te betekenen dat een meer gelijke verdeling er de *oorzaak* van is dat bedrijven beter presteren.

Een alternatieve verklaring is bijvoorbeeld de zelfselectie van vrouwen richting bedrijven die al goed presteren. Of, zoals een advocaat van de duivel zou kunnen zeggen, bedrijven die al goed presteren kunnen het zich veroorloven om vrouwen in hun raden van bestuur te hebben die geen waarde toevoegen. Onderzoek naar de *business case* voor diversiteit op basis van data uit de praktijk geeft dus waarschijnlijk een vertekend beeld van de daadwerkelijke effecten van diversiteit in teams.

Onderzoek aan de hand van experimenten in het lab wordt daarentegen niet beïnvloed door dergelijke verstorende mechanismen. Maar de overeenkomst van labexperimenten met praktijksituaties is wellicht beperkt. Dit geldt zeker voor de effecten van diversiteit. Het is niet waarschijnlijk dat de effecten van diversiteit op bijvoorbeeld de onderlinge communicatie en coördinatie op korte termijn al volledig zichtbaar worden (Boisjoly et al., 2006). Teams hebben vaak tijd nodig om een rolverdeling vast te stellen, een verstandhouding op te bouwen en een werkwijze door te voeren die bijdragen aan het bereiken van hun doel. Een experiment in het lab duurt doorgaans maximaal enkele uren.

Onderzoek naar het oorzakelijke verband tussen diversiteit en de prestaties van teams op de langere termijn blijkt een uitdagende opdracht. Hierdoor zijn studies die dit type effectmeting nastreven schaars. Veldexperimenten combineren de sterke eigenschappen van studies gebaseerd op data uit de praktijk en labexperimenten (Bardsley et al., 2010; Harrison en List, 2004). Dit proefschrift beschrijft drie van deze veldexperimenten die zijn uitgevoerd in realistische maar gecontroleerde omstandigheden.

Methode en bijdrage

In dit proefschrift worden de effecten van diversiteit in teams op hun langere termijn prestaties gemeten. Dit door het uitvoeren van veldexperimenten onder teams van eerstejaars studenten die een echt bedrijf opstarten, managen en uiteindelijk weer opheffen aan het einde van het collegejaar. Het ondernemerschapsprogramma is voor deze studenten een verplicht en belangrijk onderdeel van het eerstejaars curriculum binnen hun HBO-opleiding *international business* aan de Hogeschool van Amsterdam (HES). Deze studentbedrijven werden opgericht en ontbonden in een gecontroleerde maar realistische omgeving (van het Jong Ondernemen programma). Bovendien stond de opleiding toe om de samenstelling van teams in termen van geslacht, etniciteit en cognitieve vaardigheden te bepalen en in te delen. Zo ontstond een unieke kans om de veldexperimenten uit te voeren die in dit proefschrift worden beschreven.

Het ondernemerschapsprogramma vereist van studenten dat ze een bedrijfsactiviteit kiezen, (start)kapitaal verzamelen door de uitgifte van aandelen, managers benoemen en de taken verdelen, producten of diensten op de markt brengen, de boekhouding bijhouden en aandeelhoudersvergaderingen organiseren. Alles aan het bedrijf is echt, inclusief belastingen en afdrachten voor sociale zekerheid. Daarnaast ondervinden studenten sterke prikkels die hun eigen belangen overeen laten komen met de (financiële) prestaties van het bedrijf. De teams zijn opgebouwd uit ongeveer 10 tot 12 studenten die gezamenlijk een substantiële taak uitvoeren.

Het doel is een zo hoog mogelijke omzet, winst en aandeelhouderswaarde. In dit proefschrift worden de prestaties van teams gemeten aan de hand van deze bedrijfsresultaten. Vragenlijsten die werden afgenomen voor, tijdens en na het ondernemerschapsprogramma maken het mogelijk om verschillende onderliggende mechanismen uit de relevante literatuur te analyseren. Denk bijvoorbeeld aan complementariteiten en de mogelijkheid om van anderen te leren, maar ook aan de onderlinge communicatie en coördinatie binnen teams (Lazear, 1999). Dergelijke mechanismen kunnen een verklaring bieden waarom teams met een bepaalde mate van diversiteit anders presteren.

De drie veldexperimenten werden uitgevoerd in de collegejaren 2008-2009 en 2009-2010. In het eerste collegejaar zijn 550 studenten toegewezen aan 45 teams op basis van hun geslacht en etniciteit. Dit om variatie in de geslachts- en etnische samenstelling van teams te creëren. De opzet van beide veldexperimenten wordt niet beïnvloed door de wisselwerking tussen diversiteit in geslacht en etniciteit. Deze verhouden zich namelijk niet significant tot elkaar in de verzameling van teams onderzocht voor dit proefschrift.

In het collegejaar 2009-2010 werden teams ingedeeld op basis van hun vooraf gemeten cognitieve vaardigheden. Op de eerste dag van het collegejaar werden alle studenten onderworpen aan de veelgebruikte Raven intelligentietest. De in totaal 573 studenten werden vervolgens systematisch opgedeeld in 49 teams gebaseerd op hun score in deze test. Sommige teams hadden een grote variatie in hun cognitieve vaardigheden, andere een lage. Bovendien verschilde het gemiddelde niveau van cognitieve vaardigheden per team op systematische wijze. Zowel studenten als docenten waren niet op de hoogte van de drie veldexperimenten.

De veldexperimenten in dit proefschrift vinden plaats in een omgeving die het functioneren van teams beter nabootst dan in voorgaande labexperimenten. In vergelijking met studies op basis van data uit de praktijk biedt de experimentele opzet bovendien de mogelijkheid om een grote variatie in diversiteit tussen teams te creëren. Dit is van belang voor het analyseren van een mogelijk niet-lineaire relatie tussen de prestaties van teams en hun diversiteit. Uit dit proefschrift blijkt bijvoorbeeld dat etnische diversiteit in teams alleen een positief effect heeft als ten minste de meerderheid van het team bestaat uit leden met verschillende etnische achtergronden. Onderzoek gebaseerd op data uit de praktijk zal dit niet snel kunnen aantonen, omdat teams met een dergelijk grote mate van etnische diversiteit nog weinig voorkomen in de praktijk.

De experimentele opzet in dit proefschrift brengt ook nadelen met zich mee. Er wordt gebruik gemaakt van studenten die gezamenlijk een jaar lang een bedrijf managen in plaats van individuen die een langere periode samenwerken in een team met een mogelijk onbekende tijdshorizon. Studenten in deze veldexperimenten beschikken over vergelijkbare capaciteiten, kennis en serieuze werkervaring (die ze allen ontberen). Deze minder realistische aspecten beperken wellicht de vertaalslag naar teams in organisaties.

Een tweede nadeel van de experimentele opzet is dat niet kan worden gemeten of de onderliggende mechanismen daadwerkelijk een verklaring bieden voor verschillen in prestaties tussen teams. Zo kunnen teams met een grote mate van diversiteit bijvoorbeeld worden gekenmerkt door leden die elkaar beter aanvullen en meer leren van hun mede-teamleden. Maar dit sluit niet uit dat de prestaties van teams ook een omgekeerd effect kunnen hebben op deze mechanismen. Daardoor biedt dit proefschrift (net zoals andere studies) hoogstens aanwijzingen waarom teams met een verschillende mate van diversiteit anders presteren.

Ondanks de hierboven besproken minder realistische aspecten van de teams in dit proefschrift, hebben de veldexperimenten toch een zekere mate van externe validiteit. Studenten voeren een substantiëlë set aan gezamenlijke taken uit. Deze taken zijn breed en complex. Ze vereisen dat studenten binnen teams onderling een rolverdeling vaststellen, een verstandhouding opbouwen en een werkwijze doorvoeren op een vergelijkbare manier als bij managers of werknemers die samenwerken in teams. Voorbeelden van teams in organisaties met vergelijkbaar ruime en complexe taken zijn *business units*, projectteams en raden van bestuur. Het is dan ook waarschijnlijk dat de veldexperimenten in dit proefschrift een enigszins getrouwe weergave vormen van het functioneren van teams in de praktijk. Dit proefschrift biedt hiermee inzicht in de effecten van diversiteit op de prestaties van teams in organisaties.

Overzicht

Hoofdstuk 2 geeft een beschrijving van de context en opzet van de veldexperimenten. Ook gaat dit hoofdstuk in op de prikkels die teamleden ervaren om hun eigen belangen op een lijn te brengen met de belangen van het team.

Hoofdstuk 3 onderzoekt de effecten van diversiteit in geslacht op de prestaties van teams. Dit hoofdstuk is gebaseerd op het artikel *'The impact of gender diversity on the performance of business teams: Evidence from a field experiment'* met als co-auteurs Hessel Oosterbeek en Mirjam van Praag (te verschijnen in *Management Science*, 2013). Dit hoofdstuk is van belang voor de eerder besproken discussie rondom de ondervertegenwoordiging van vrouwen in raden van bestuur. Circa 15 procent van de bestuursfuncties in bedrijven uit de *Fortune 500* werd in 2010 bekleed door vrouwen. Dit percentage bedroeg ongeveer 10 procent voor de top 300 van Europese bedrijven. Onder Nederlandse beursgenoteerde bedrijven lag dit percentage zelfs nog lager (Catalyst, 2010; Woods, 2010).

Daarnaast is het aandeel vrouwen in het hoger onderwijs en op de arbeidsmarkt toegenomen. Dit vergroot de beleidsrelevantie van inzicht in de effecten van een meer gelijke verdeling tussen mannen en vrouwen in raden van bestuur. Sommige landen, waaronder Nederland, dwingen (binnenkort) zelfs een grotere vertegenwoordiging van vrouwen af door minimumquota voor vrouwelijke bestuurders op te leggen. Tot op heden is er weinig bewijs dat dergelijke regelgeving bijdraagt aan betere prestaties van bedrijven of teams. De resultaten van het veldexperiment in Hoofdstuk 3 laten zien dat gemengde teams beter presteren dan teams die vooral zijn opgebouwd uit mannen. Daarnaast lijkt het erop dat gemende teams ook beter presteren dan teams met een meerderheid van vrouwen. Er zijn in de data geen aanwijzingen gevonden waarom teams met een gelijke verdeling van mannen en vrouwen beter presteren.

Hoofdstuk 4 behandelt de effecten van etnische diversiteit in teams op hun prestaties. Dit hoofdstuk is gebaseerd op een artikel met Mirjam van Praag getiteld *'Ethnic diversity and team performance: A field experiment'*. Etnische diversiteit is van groot belang in een globaliserende wereld en het aandeel van etnische minderheden in Westerse samenlevingen neemt sterk toe. Daardoor is het waarschijnlijk dat teams de komende jaren meer divers zullen worden in termen van etniciteit (Alesina en La Ferrara, 2005). Voor de Verenigde Staten bijvoorbeeld is de verwachting dat het aandeel van etnische minderheden toeneemt van ongeveer eenderde op dit moment naar een meerderheid in 2042 (Bernstein en Edwards, 2008).

Het veldexperiment in Hoofdstuk 4 biedt een realistische vooruitblik op de effecten van een grote mate van etnische diversiteit op de prestaties van teams. Uit de resultaten blijkt dat een gematigde etnische diversiteit in teams geen effect heeft op hun prestaties. Etnische diversiteit heeft een positief effect op de prestaties van teams als ten minste de meerderheid van het team verschillende etnische achtergronden heeft. In lijn met theoretische voorspellingen uit de literatuur kan dit positieve effect worden verklaard doordat teamleden in etnisch diverse teams elkaars kennis beter aanvullen. Hierdoor leren ze mogelijk ook meer van hun mede-teamleden.

Hoofdstuk 5 bestudeert de effecten van diversiteit in cognitieve vaardigheden op de prestaties van teams. Dit hoofdstuk is gebaseerd op het artikel *'Ability dispersion and team performance: A field experiment'* met als co-auteurs Simon Parker en Mirjam van Praag. Cognitieve vaardigheden zijn sterk van invloed op het gedrag van individuen en de beslissingen die ze nemen (Cutler en Lleras-Muney, 2010; Dohmen et al., 2010; Grinblatt et al., 2011). Hiermee bepalen cognitieve vaardigheden in grote mate individuele uitkomsten zoals inkomen, maar op geaggregeerd niveau bijvoorbeeld ook economische groei (Hanushek en Woessmann, 2008). Er wordt doorgaans breed verondersteld dat cognitieve vaardigheden de prestaties van teams beïnvloeden. Maar de precieze effecten van diversiteit in cognitieve vaardigheden op de prestaties van teams blijven vaak onduidelijk (Hamilton et al., 2012).

In Hoofdstuk 5 wordt een theoretisch model ontwikkeld. In dit model levert een grotere mate van diversiteit in cognitieve vaardigheden meer kennis op voor een team (en dus betere prestaties). Tegelijkertijd wordt monitoring van elkaars inzet steeds moeilijker naarmate de teamleden meer van elkaar verschillen in hun cognitieve vaardigheden. De effectiviteit van monitoring om gebrek aan inzet van teamleden waar te nemen en te bestraffen neemt hiermee af. Op een gegeven moment neemt de kans om ontdekt en ontslagen te worden bij gebrek aan inzet zover af dat dit ten koste gaat van de prestaties van teams. De resultaten van dit veldexperiment wijzen uit dat de prestaties van teams eerst toenemen en daarna afnemen door een toename van diversiteit in cognitieve vaardigheden. Teams met een gematigde diversiteit in cognitieve vaardigheden presteren dus het best. Dit is in lijn met de voorspellingen uit het model. In teams met een gematigde diversiteit in cognitieve vaardigheden vinden daarnaast minder ontslagen plaats. Maar dit hoeft niet te betekenen dat dit mechanisme ook verklaart waarom deze teams beter presteren.

Hoofdstuk 6 geeft een beknopte samenvatting van de bevindingen uit de voorgaande hoofdstukken en bevat enkele afsluitende opmerkingen, inclusief mogelijkheden voor vervolgonderzoek.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

504 J.J. DE HOOP, *Keeping Kids in School: Cash Transfers and Selective Education in Malawi*

505 O. SOKOLINSKIY, *Essays on Financial Risk: Forecasts and Investor Perceptions*

506 T. KISELEVA, *Structural Analysis of Complex Ecological Economic Optimal Management Problems*

507 U. KILINC, *Essays on Firm Dynamics, Competition and Productivity*

508 M.J.L. DE HEIDE, *R&D, Innovation and the Policy Mix*

509 F. DE VOR, *The Impact and Performance of Industrial Sites: Evidence from the Netherlands*

510 J.A. NON, *Do ut Des: Incentives, Reciprocity and Organizational Performance*

511 S.J.J. KONIJN, *Empirical Studies on Credit Risk*

512 H. VRIJBURG, *Enhanced Cooperation in Corporate Taxation*

513 P. ZEPPINI, *Behavioural Models of Technological Change*

514 P.H. STEFFENS, *It's Communication, Stupid! Essays on Communication, Reputation and (Committee) Decision-Making*

515 K.C. YU, *Essays on Executive Compensation - Managerial Incentives and Disincentives*

516 P. EXTERKATE, *Of Needles and Haystacks: Novel Techniques for Data-Rich Economic Forecasting*

517 M. TYSZLER, *Political Economics in the Laboratory*

518 Z. WOLF, *Aggregate Productivity Growth under the Microscope*

519 M.K. KIRCHNER, *Fiscal Policy and the Business Cycle - The Impact of Government Expenditures, Public Debt and Sovereign Risk on Macroeconomic Fluctuations*

520 P.R. KOSTER, *The Cost of Travel Time Variability for Air and Car Travelers*

521 Y. ZU, *Essays of Nonparametric Econometrics of Stochastic Volatility*

522 B. KAYNAR, *Rare Event Simulation Techniques for Stochastic Design Problems in Markovian Setting*

523 P. JANUS, *Developments in Measuring and Modeling Financial Volatility*

524 F.P.W. SCHILDER, *Essays on the Economics of Housing Subsidies*

525 S.M. MOGHAYER, *Bifurcations of Indifference Points in Discrete Time Optimal Control Problems*

526 C. CAKMAKLI, *Exploiting Common Features in Macroeconomic and Financial Data*

527 J. LINDE, *Experimenting with New Combinations of Old Ideas*

528 D. MASSARO, *Bounded Rationality and Heterogeneous Expectations in Macroeconomics*

529 J. GILLET, *Groups in Economics*

530 R. LEGERSTEE, *Evaluating Econometric Models and Expert Intuition*

531 M.R.C. BERSEM, *Essays on the Political Economy of Finance*

532 T. WILLEMS, *Essays on Optimal Experimentation*

533 Z. GAO, *Essays on Empirical Likelihood in Economics*

534 J. SWART, *Natural Resources and the Environment: Implications for Economic Development and International Relations*

535 A. KOTHIYAL, *Subjective Probability and Ambiguity*

536 B. VOOGT, *Essays on Consumer Search and Dynamic Committees*

537 T. DE HAAN, *Strategic Communication: Theory and Experiment*

538 T. BUSER, *Essays in Behavioural Economics*

539 J.A. ROSERO MONCAYO, *On the Importance of Families and Public Policies for Child Development Outcomes*

540 E. ERDOGAN CIFTCI, *Health Perceptions and Labor Force Participation of Older Workers*

541 T. WANG, *Essays on Empirical Market Microstructure*

542 T. BAO, *Experiments on Heterogeneous Expectations and Switching Behavior*

543 S.D. LANSDORP, *On Risks and Opportunities in Financial Markets*

544 N. MOES, *Cooperative Decision Making in River Water Allocation Problems*

545 P. STAKENAS, *Fractional Integration and Cointegration in Financial Time Series*

546 M. SCHARTH, *Essays on Monte Carlo Methods for State Space Models*

547 J. ZENHORST, *Macroeconomic Perspectives on the Equity Premium Puzzle*

548 B. PELLOUX, *The Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies*

549 N. YANG, *Markov-Perfect Industry Dynamics: Theory, Computation and Applications*

550 R.R. VAN VELDHUIZEN, *Essays in Experimental Economics*

551 X. ZHANG, *Modeling Time Variation in Systemic Risk*

552 H.R.A. KOSTER, *The Internal Structure of Cities: The Economics of Agglomeration, Amenities and Accessibility*

553 S.P.T. GROOT, *Agglomeration, Globalization and Regional Labor Markets: Micro Evidence for the Netherlands*

554 J.L. MOHLMANN, *Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products*

This dissertation studies the impact of diversity on team performance using a series of field experiments in which teams start up and manage real companies under identical circumstances. Exogenous variation in - otherwise random - team composition is imposed by assigning individuals to teams based on their gender, ethnicity or cognitive ability. The setting of these field experiments closely resembles that of (business) management practices in the longer run where tasks are diverse and involve complex decision-making. Evidence from this kind of experiments potentially contributes to the effective composition of teams in organizations. The results of this dissertation demonstrate that diversity in gender, ethnicity and cognitive ability has substantial and non-monotonic effects on the performance of teams. Various underlying mechanisms are explored to explain why teams of distinct degrees of diversity perform differently.

Sander Hoogendoorn (1984) obtained his Master's degree in Economics and Social Sciences at Utrecht University in 2006, and in Business Economics at the University of Amsterdam in 2007. After his graduation, he worked as manager and researcher at the Amsterdam Center for Entrepreneurship. Since 2010 he has been affiliated with the Tinbergen Institute and focused on completing his PhD in Economics at the University of Amsterdam.