Fakultät für Wirtschaftswissenschaften
Lehrstuhl für Organisations-, Medien- und Sportökonomie

THE ECONOMIC AND SOCIAL DETERMINANTS OF EMPLOYEE BEHAVIOR: EVIDENCE FROM INSIDER ECONOMETRIC STUDIES

Der Fakultät für Wirtschaftswissenschaften der

Universität Paderborn

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

- Doctor rerum politicarum -

vorgelegte Dissertation

von

Konstantin Böddeker, M.Sc.

geboren am 10.12.1985 in Paderborn

(2016)

The data used in this thesis are proprietary and can, therefore, not be made available to other researchers.

The results, opinions and conclusions of this dissertation are those of the author and not necessarily those of the Volkswagen AG.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

APPENDIX

List of Abbreviations

CI	Confidence Interval
CV	Coefficient of Variation
DWD	Deutscher Wetterdienst (German Meteorological Service)
ES	Effect Size
ESP	Spain
EU	European Union
FE	Fixed-Effects (Regression)
GDP	Gross Domestic Product
GER	Germany
GPS	Global Positioning System
HR	Human Resources
HRM	Human Resource Management
ILI	Influenza-like Illnesses
KPI	Key Performance Indicator
OECD	Organization for Economic Co-Operation and Development
OHC	Occupational Health Care
OLS	Ordinary Least Squares (Regression)
R&D	Research and Development
RE	Random-Effects (Regression)
ROA	Return on Assets
ROE	Return on Equity
ROS	Return on Sales
s.d.	Standard Deviation
SUR	Seemingly Unrelated Regression
TMT	Top Management Team
UK	United Kingdom
US	United States of America
WHO	World Health Organization

1 INTRODUCTION

Since the 1980s the strand of personnel economics has emerged from labor economics in order to address both the lack of economic perspectives in early human resource management research as well as the growing need for hands-on, business-oriented insights on human resource management (HRM) (Lazear 2000a). Initially influenced by informational economics and further extended by ideas originating from social psychology and organizational sociology, personnel economics is nowadays a well-established and substantial part of labor economics research. Reviews by Prendergast (1999), Lazear (1999, 2000a) and Lazear and Oyer (2013) provide a comprehensive overview of the historical development as well as the contemporary debate within the field of personnel economics.

Personnel economics applies (labor) economic methods such as econometrics and game theory to provide a detailed understanding of the functioning of a firm and its human resource practices. However, due to limited data availability in the early years, personnel economics started off with mostly theoretical considerations. Later contributions combine theory with empirical analyses as they are built on rich firm-based data sets. In general, personnel economic models and theories play by the traditional rules of economics, in particular the assumptions of rational maximizing agents, specific equilibrium analysis of labor and product markets and efficiency as a result of market equilibria (Lazear 2000a, Lazear, Shaw 2007). Nevertheless, the main research fields of personnel economics are not limited to the classic HRM issues such as compensation, turnover and incentives. Instead, its scope of research is extended beyond traditional economic fields, including topics such as group norms, teamwork settings, or worker empowerment. In brief, one can summarize the main research interests of personnel economists to comprise incentives, worker-firm matching, compensation, skill development and organization of work (Lazear, Oyer 2013). These topics share the common thread in that personnel economics is all about enhancing worker productivity (Lazear, Gibbs 2009).

In contrast to labor economics, personnel economics does not appear policy-oriented: It focuses exclusively on the welfare within an individual employment relationship and not on the overall social welfare (Lazear, Oyer 2013). In other words, personnel economics aims at offering practical implications for managers on how to improve operations and, thus, enhance the overall productivity of a given firm. As a consequence, personnel economics research mainly focuses on those variables that managers have an actual bearing in.

Within the field of personnel economics, the strand of insider econometrics has evolved quite recently (Ichniowski, Shaw 2009). Inspired by seminal papers by Ichniowski, Shaw and Prennushi (1997) as well as Lazear (2000b), more and more researchers pursue the insider econometric approach in order to address the question whether particular management practices raise a firm's productivity. Productivity in insider studies is usually assessed through various variables, including product quality, production line downtime, worker absenteeism, worker turnover and speed of order fulfillment (Ichniowski, Shaw 2013). Moreover, researchers try to identify the underlying mechanisms and employee behavior that trigger these productivity effects. Both mechanisms and behavior may vary across workers, work groups, firms, industries, or with other environmental influences such as the overall set of management practices that are in operation at a given firm.

In general, insider studies rely on micro-level (panel) datasets that originate from only one or a few companies or industries. This rich data allows for profound analyses of productivity effects of particular management practices that are in operation at the given firms or industry (e.g. Bartel, Ichniowski, Shaw 2004, Ichniowski, Shaw 2013). Insider studies are often assumed to be advantageous as they combine state-of-the-art empirical analyses of high-quality data with additional expert knowledge stemming from inside the company. This two-folded research strategy usually facilitates the formulation of hypotheses as well as the interpretation of empirical results and might reveal additional evidence that would have been undetected without inside knowledge (Ichniowski, Shaw 2013). In what follows, the most important contributions that particularly well demonstrate the great power and wide applicability of insider econometrics are briefly presented.

In their early work, Ichniowski, Shaw and Prennushi (1997) study the impact of innovative HRM practices such as teamwork, incentive pay or on-the-job training on productivity. Using panel data from 36 steel finishing lines at US steel mills, they find substantial productivity gains associated with the implementation of clusters of innovative HRM practices. Later, Lazear (2000b) analyzes behavioral responses to the introduction of piece rate pay at a large US-based windshield installer. As the company shifts from fixed hourly pay to piece rate pay, he reports an increase in productivity of 44% that is related to incentive as well as sorting effects.

Since then, the quantity of insider studies has increased constantly. For instance, Hamilton, Nickerson and Owan (2003) study the introduction of autonomous work teams and team-based piece rate schemes at a US garment producer. The joint adoption of teamwork and team incentives results in average productivity gains of 14%. In a series of papers, Bandiera, Barankay and Rasul (2005, 2007, 2009, 2010, and 2011) examine the productivity of workers and managers in a UK fruit picking farm under varying remuneration schemes. For instance, they find worker productivity to be at least 50% higher under piece rate remuneration than under relative incentives since workers appear to mitigate negative externalities of their own work – in terms of wage penalties for others – when working with their friends. Furthermore, the most capable workers are willing to waive about 10% of their potential earnings to allow for social interaction with their less capable co-workers. Likewise, less capable workers increase their productivity by about 10%. For managers, the results indicate that supervisors favor workers who they are socially connected with under fixed wage schemes while favoring the most productive workers when their own remuneration is based on the workers' performance.

Peer effects in the workplace are observed by Mas and Moretti (2009) based on innovative high-frequency data from supermarket checkout desks. They observe positive productivity spillovers associated with the mere presence of more productive co-workers. More recently, Bloom et al. (2013) study the productivity effects of modern management practices such as performance-based incentives in the context of Indian textile companies. They report productivity gains of 11% at those companies that introduced modern management practices. For a comprehensive overview of insider econometrics see the review articles by Lazear and Shaw (2007), Shaw (2009), Bloom and van Reenen (2011) as well as Ichniowski and Shaw (2013). Assessing the future of insider econometrics in the digital age, one might assume this approach to highly benefit from technological advances such as big data and high-frequency data mining. These data sources offer innovative research potentials due to easy accessible computerized data that meets all requirements of high-quality research. Two studies of this dissertation already profit from high-frequency computerized data gathered via a modern GPS-based fleet management system operated by a truck hauling company (chapters four and five).

As stated above, personnel economics predominantly focuses on productivity enhancement with productivity being understood as the overall outcome of employee behavior. In this, it is widely recognized among personnel economists that human behavior is affected by both psychological preferences as well as environmental influences such as budgets, constraints, incentives or social interaction. The study of the effects of these environmental influences on human behavior and, thus, worker productivity is the core of personnel economics research (Lazear, Gibbs 2009). Following the tradition of insider econometrics, this dissertation seeks to analyze how economic and social influences determine employee behavior in order to better understand the link between particular economic and social variables and worker productivity.[1] Methods and instruments that may enhance worker productivity are discussed in each study based on the respecting findings. Throughout this thesis, productivity is assessed by means of divergent performance measures such as overall team performance (chapter two),[2] employee absenteeism (chapter three) and several objective computerized performance evaluations (chapters four and five). Given the main personnel economics research fields assessed by Lazear and Oyer (2013), the work at hand focuses on incentives and the organization of work in teams.

It has long been recognized in personnel economics that moral hazard problems exist within the employment relationship since it usually represents a classic principal-agent setting.[3] Due to information asymmetries, rent-seeking agents can be assumed to choose levels of effort below the ones agreed upon with the principal in order to maximize their own utility.[4] In this situation, incentives may be introduced to align interests of principals and agents and, thus, overcome agency problems. A large theoretical literature has emerged on the necessity for and the design of incentives that motivate employees to behave in the very interest of the firm, e.g. Holmström and Milgrom (1991) as well as Baker (1992) on multitasking agents that chose to perform only those activities providing the highest incentives; Milgrom (1988) and Milgrom and Roberts (1988) on agents' rent-seeking activities with regard to subjective evaluations by supervisors; Lazear (1979) on deferred compensation; Fama (1980) as well as Holmström (1982) on career concerns. Comprehensive reviews of the most relevant contributions on incentives in principal-agent settings are provided by Gibbons (1997) and Prendergast (1999).

[1] Chapter two depicts an exception as it follows the meta-analytic approach in analyzing correlations observed in previous empirical research.

[2] In this meta-analysis overall team performance is based on subjective performance ratings – either self-assessed by the team or externally-assessed by managers and supervisors – as well as objective financial key performance indicators (KPIs) and efficiency measures such as productivity, profitability and goal achievement.

[3] For a general overview of agency theory see, for instance, Oyer and Schaefer (2011).

[4] However, not all agents are dishonest. Bloom et al. (2015) observe that working at home offices increases productivity although monitoring decreases. In contrast, other researchers find agents to display shirking behavior in various settings such as auto repairs (Schneider 20012), restaurant tipping (Azar, Yosef, Bar-Eli 2015), or taxi transportation (Balafoutas et al. 2013). In a recent meta-analysis, Rosenbaum, Billinger and Stieglitz (2014) observe people to be either unconditional cheaters/non-cheaters or decide on being honest or dishonest according to the given environmental influences such as monitoring activities.

Empirical evidence on the benefits of incentives is broad and will be thoroughly discussed in the respective chapters. However, despite the extensive empirical support for their advantageousness, extrinsic incentives may also induce the very behavior they intend to prevent. In other words, extrinsic incentives (e.g. pay for performance schemes) literally buy off (or crowd out) intrinsic incentives such as values or beliefs. Since these intrinsic incentives are equally important in determining employees' effort choice decisions, overall worker productivity may be reduced (e.g. Deci 1971, Bénabou, Tirole 2003). Empirical evidence on this so-called crowding-out effect will be discussed in the respective chapters.

Apart from incentives, social peer influences can, to a large extent, explain employees' effort choice decisions. In the context of teamwork, personnel economics theory suggests team-based productivity gains to occur mainly for two reasons. First, according to Lazear (1999), broad sets of complementary knowledge, skills and expertise constitute the main advantage of teamwork over individual work. Second, due to peer effects in terms of productivity spillovers high-productive employees may positively affect the productivity of their peers. Falk and Ichino (2006) experimentally confirmed the existence of peer effects. Empirical studies such as the above-mentioned Mas and Moretti paper (2009) on supermarket checkout clerks provide further evidence. However, the work at hand does not focus on whether or not teamwork is advantageous compared to individual work. Instead, the objective here is to understand the influence of composition effects on group outcomes with respect to work group diversity. Two opposing research traditions have emerged over time. On the one hand, social categorization theory (Tajfel 1982, Turner 1987) and similarity-attraction paradigm (Byrne 1969, 1971) assume social interactions to be facilitated as well as more desirable among similar peers. Therefore, homogeneous teams seem to be preferable over heterogeneous teams. On the other hand, information processing perspective (Cox, Blake 1991, Cox 1993) argues that teams gain from broader sets of knowledge and skills. In other words, heterogeneous teams are expected to outperform homogenous teams. The latter argument is supported by Lazear's considerations (1999) as stated above. Empirical evidence for both theoretical approaches is discussed in the respective chapters.

The influence of incentives as well as teamwork on employee behavior and, thus, worker productivity are addressed in the dissertation at hand based on four separate analyses presented in chapters two to five. The theoretical background and empirical evidence are discussed in more detail there. The fundamental research question of each chapter can be stated as follows:

Chapter 2: Is employee behavior and, thus, work group performance affected by work group composition in terms of homogeneity/heterogeneity?

Chapter 3: What are potential social and economic determinants of employee absence behavior at the team-level?

Chapter 4: Do employees respond to unfavorable changes in extrinsic incentive design by adapting their behavior in terms of effort choice and productivity?

Chapter 5: Do employees adjust their effort choice and productivity when being "promoted" from temporary to permanent employment?

Along these research questions, the four studies presented throughout the work at hand can be divided into two parts. While the first part (chapter two and three) analyzes the effect of social determinants within work groups (peer effects and worker heterogeneity) on workers' effort choice decisions, the second part (chapter three to five) focuses on the relationship between different types of economic determinants (incentives) and workers' effort choice decisions.

Chapter two follows a classic meta-analytic approach and, therefore, is based on existing empirical studies. The meta-analytic approach was chosen to account for the vast amount of research on group diversity and the often small samples this research is built on. In order not to provide just another empirical study on diversity, a systematic quantitative summary of existing findings is warranted. In contrast, chapters three to five build on insider econometric data derived from a large European automobile manufacturer. This unique insider data had hitherto been unavailable for empirical analyses. All data sets used throughout this thesis are either compiled by the author (chapters two, four and five) or by company representatives according to predefined standards set by the author (chapter three). Exact variable definitions, continuous monitoring during data provision and thorough data checks helped to guarantee high-quality data that meets all necessary requirements concerning reliability as well as validity.

Since this dissertation follows a personnel economics approach, all analyses conform to its main assumptions and methodologies. In general, one can distinguish four major "*building blocks*" (Lazear, Shaw 2007, p. 91) of personnel economics (Lazear 2000a, Lazear, Shaw 2007). First, both firms and employees are assumed to be rational maximizing agents. Employees seek to maximize their individual utility while firms seek to maximize their profit under given economic constraints such as imperfect information or transaction costs. Personnel economics is built on the assumption that agents' behavior is determined mainly by the interaction of agents and not by other influences beyond the agents' control. As a consequence, any decision-making of both employees and firms, e.g. in terms of effort choice decisions or the choice of particular management practices, is based on maximizing behavior. Second, personnel economics models are based on equilibrium theory that assumes labor and product markets to be competitive since both firms and workers are considering the actions of other agents when making their own decisions. This notion results in a specific price-quantity equilibrium that allows predicting a model's outcomes in the real world quite accurately. Third, a core of personnel economics research is the analysis of efficiencies resulting from these equilibria. When interacting with a profit-maximizing firm a utility-maximizing employee usually chooses a behavior that in the end will make both parties better off. Yet, in some situations efficiency is forfeited due to (market) inefficiencies, e.g. in terms of moral hazard. Here, it is the very aim of personnel economists to identify actions that help workers and firms alike to alleviate inefficiencies. Fourth, econometric and experimental designs are the essentials of a personnel economist's toolbox in order to pursue his research. On this account, the research methods applied in the work at hand have been customized to the peculiarities of each of the four studies when it comes to data and research focus. Throughout this thesis state-of-the-art econometric methods are applied including OLS regression, random- and fixed-effects models for longitudinal data, Seemingly Unrelated Regression (SUR) models as introduced by Zellner (1962) as well as logit and probit estimations. The meta-analysis of chapter two follows the Hedges and Olkin (1985) meta-analysis approach while moderator analysis is based on meta-regressions as proposed by DerSimonian and Laird (1986). All methods allow for an in-depth analysis of the underlying research questions. The methodological approach will be discussed separately in each chapter.

This thesis is based on four separate econometric studies to be published in peer-reviewed academic journals in the fields of management science and personnel economics.[5] The remainder of this thesis is organized as follows. Chapter two addresses the ambivalent nature of work group diversity – often referred to as a "*double-edged sword*" (Milliken, Martins 1996, p. 403) – that is emphasized by two competing theoretical strands. On the one

[5] As a consequence, redundancies concerning literature, data set description and econometric methodology discussion may occur at some instances.

hand, similarity-attraction paradigm and social categorization theory suggest group heterogeneity to induce performance and productivity to be inferior due to sub-group formation (in-group vs. out-group) and work group conflict. As a consequence, homogeneous teams are assumed to be preferable (Tajfel 1982, Tajfel, Turner 1986, Turner 1987). On the other hand, the information processing perspective assumes heterogeneous teams to increase performance and productivity due to complementary cognitive resources, experiences and network ties. Hence, heterogeneous teams are expected to be advantageous in terms of performance and productivity (Cox, Blake 1991). There is extensive empirical evidence for both approaches. Applying meta-analytic research to a total of 66 individual samples derived from 63 primary studies on work group diversity, this paper seeks to analyze the broad literature on diversity in search of common statistical patterns on the advantageousness of either homogeneous or heterogeneous teams. Diversity is considered along less task-related (age, gender, culture) and highly task-related (tenure, function, educational background and educational level) attributes. Furthermore, empirical tests for potential moderating effects of team type (top management teams, work teams, research and development teams and mixed teams) and team size are performed. Overall, findings confirm the ambivalent nature of work group heterogeneity. In support of social categorization theory and similarity-attraction paradigm, the results suggest negative population correlations of gender and age heterogeneity on team performance. Likewise, educational background diversity is observed to be positively related to team performance, thus, confirming arguments stated by information processing perspective. Two moderating effects are observed. First, increasing team size moderates the influence of team diversity on performance negatively for less task-related diversity attributes and positively for highly task-related attributes. Second, top management team type and research and development team type positively moderate the diversity-performance relationship regardless of the task-relatedness of diversity. Moreover, work team type negatively moderates the diversity-performance relationship for work teams regardless of the task-relatedness of diversity attributes.

Chapter three depicts joint work with Bernd Frick in which a hitherto unavailable data set covering 160 blue-collar work units from four international manufacturing plants of a large European automobile manufacturer is used to analyze social and economic determinants of employee absenteeism. In this chapter, the understanding of absenteeism is based on the Steers and Rhodes (1978) process model as well as on economic theories as proposed by neoclassical labor supply models (e.g. Allen 1981, Dunn, Youngblood 1986) and efficiency wage theory (Shapiro, Stiglitz 1984). By means of a comprehensive review of existing empirical findings on absence research, a set of thirteen potential determinants of absence is identified that stem from social peer influences (work unit size, turnover, share of temporary co-worker, share of health-impaired co-workers), economic influences / incentives (sick pay regulations, employment protection laws, unemployment, prosperity level), worker characteristics (age, gender, tenure, acute health) and working conditions (shift system). Absence is found to be a multifaceted phenomenon since only few statistical patterns hold true for all four plants. Instead, core findings suggest results to vary significantly by production site. Based on empirical analyses as well as interviews with on-site experts (i.e. HR managers, worker representatives, managers and shop floor staff), results suggest absence to increase with unit size and overall unit turnover. Likewise, a high share of temporary co-workers increases group-level absence among permanent employees. In contrast, the hypothesized link between absence and the share of co-workers that suffer any kind of permanent or temporary health-impairment could not be supported. Additionally, positive relationships between absence and strict employment protection laws as well as favorable prosperity levels are identified. Findings with respect to the national (un)employment situation are inconclusive and results for national sickness benefits legislation fail to reach statistical significance. Furthermore, group-level absence is observed to increase with the share of female employees in the team. Moreover, both involuntary and voluntary absence is observed to increase with age. However, results for tenure as well as national influenza outbreaks are inconclusive. With regard to working conditions, absence is observed to be higher in units that operate in two- and three-shift systems compared to those following the standard one-shift system.

Employees' behavioral responses to changes in monetary incentive design are discussed in chapter four, which is joint work with Bernd Frick. Built on fleet management panel data of an in-house hauler of a large European truck manufacturer, this paper analyzes the performance of 37 commercial truck drivers on a total of 6,326 individual trips within a three-year time frame. Performance is measured via four different sets of variables that are recorded by GPS-based on-board computers: overall fuel consumption, driving behavior as reflected by several driving parameters, relative performance based on driving scores and, eventually, a computed overall performance evaluation. The organizational setting follows a quasi-experimental approach since the existing performance bonus system had been abolished by the hauler after the first two years of the three-year observation period. From theory, two potential behavioral responses might be expected from employees. Classic contract theory assumes human beings to directly respond to extrinsic incentives in increasing effort as well as reducing shirking behavior (e.g. Lazear 2000b, Gibbons 1998, Gibbons, Roberts 2013). However, a competing theoretical approach doubts these benefits of extrinsic (monetary) incentives arguing that extrinsic incentives literally "buy off" an employee's intrinsic motivation and, therefore, result in decreasing overall performance (e.g. Deci 1971, Bénabou, Tirole 2003, 2006, Gneezy et al. 2011). Empirical and experimental evidence for both approaches is discussed in the chapter. Interestingly, findings contradict economic theory in giving support for the crowding-out approach. Truck drivers are observed to display significantly lower effort and performance when incentivized extrinsically by means of a performance bonus. Estimations reveal that drivers behave less eco-friendly and less in accordance with company standards yielding an overall lower performance evaluation. This result implies that drivers' intrinsic and social motives – e.g. image concerns, environmental believes – are crowded out by monetary rewards.

As mentioned earlier, chapter five – again co-authored by Bernd Frick – analyzes behavioral responses of employees who are "promoted" from temporary to permanent contracts. This chapter is based on the identical fleet management panel data used in chapter 4. Again, the organizational setting follows a quasi-experimental design since it observes eight commercial truck drivers on 1,299 trips whose contract status changed from temporary to permanent. Considering advantages of permanent over temporary contracts (e.g. better working conditions, higher pay and stricter employment protection), temporary employees may be assumed to be highly incentivized to display high effort in order to qualify for a permanent contract. Empirical support for this argument is discussed within the chapter (e.g. Bradley et al. 2012). Findings suggest about half of the drivers to significantly increase fuel consumption after having signed a permanent contract with a strong initial effect and a long decline back to "normal" levels. This result indicates these drivers to behave as rational cheaters (Nagin et al. 2002). The same holds true for computerized performance evaluations.

Subsequently, chapter 6 summarizes the main results of this thesis and provides both general conclusions as well as suggestions for future research.

2 WORK TEAM DIVERSITY AND THE OPTIMAL COMPOSITION OF TEAMS: A META-ANALYTIC APPROACH

2.1 Introduction

These days, organizations increasingly implement teamwork to meet the demands of sustainable developments in economic, technological and societal environments (e.g. Devine et al. 1999). Given the growing importance of teams, management scholars have eagerly taken on the study of teamwork and its influence on business processes as well as employee performance (e.g. Sundstrom et al. 2000). As a consequence, teamwork research is nowadays established as one of the most relevant topics in management research, a position anecdotally described by Lazear (1999) who states that "*it is impossible to pick up a business publication these days without reading about the wonders of teamwork*" (Lazear 1999, p. C15).

At the same time, the composition of the global workforce has changed in that worker heterogeneity has increased dramatically over the last decades. Multifaceted social and demographic developments may be held accountable for this phenomenon. Consider, for instance, the constantly growing proportion of women participating in the labor market (e.g. Ali, Kulik, Metz 2011). At the same time, women increasingly advance into managerial ranks (e.g. Elsass, Graves 1997) and into jobs formerly known to be all-male professions such as medicine or law (e.g. Norgren 2010). Furthermore, the individual working life span has prolonged since employees enter the labor market at a younger age and stay much longer before becoming eligible for retirement (e.g. Grund, Westergaard-Nielsen 2008). Eventually, cultural and ethnic background heterogeneity is evermore increasing in line with global migration and human capital mobility (e.g. Artuc et al. 2015). So far, there are no clear indications that the ongoing social and demographic trends will slow down or even reverse in the near future (e.g. Deadrick, Stone 2009, Tsui, Gutek 1999).

Given the development towards a more heterogeneous workforce, it is of growing importance to understand how diversity issues affect processes and productivity within organizations and teams. According to Barsade et al. (2000), research on the "*costs and benefits of diversity in the workplace has been going on at a vigorous pace over the last two decades and more*" (Barsade et al. 2000, p. 802). As a consequence, the amount of studies on diversity approximately doubles every five years, for instance scholars published 134 studies in 2003 compared to only 19 studies back in 1988 (Harrison, Klein 2007). However, the existing research on diversity in the workplace is confusing and disappointing in a way that "*cumulative findings about the consequences of within-unit differences have been weak, inconsistent, or both*" (Harrison, Klein 2007, p. 1199).

Diversity research is mainly guided by two opposing research traditions that clearly demonstrate the ambivalent nature of diversity that is often considered a "*double-edged sword*" (Milliken, Martins 1996, p. 403). On the one hand, a substantial body of the literature states that within teams heterogeneity may induce social categorization that in turn implies sub-group formation (in-group vs. out-group) and team conflict and, as a consequence, result in lower team performance. Hence, homogeneity in teams is argued to be preferable to heterogeneity with regard to team performance. The most well-known contributions to this concept are expressed in similarity-attraction paradigm as well as social categorization theory (e.g. Tajfel 1982, Tajfel, Turner 1986, Turner 1987). On the other hand, scholars suppose that team production may benefit from heterogeneous cognitive abilities, experiences, skills and network ties of team members. These may be complements and, therefore, constitute a broad set of resources available to the team. As a result, heterogeneity in teams is reasoned to be advantageous over homogeneity regarding team performance. These considerations are summarized in information processing perspective (e.g. Hoffman, Maier 1961, Cox, Blake 1991) and economic theory alike (e.g. Lazear 1999).

In this paper, we consider both theoretical approaches to derive our main research hypotheses on the diversity-performance relationship within organizational work groups. We aim at contributing new insights to the wide field of diversity research by means of a comprehensive meta-analysis. Moreover, we test for two potential moderating effects on the diversity-performance relationship – team size and team type. Both have been identified as being influential by previous research (e.g. Stewart 2006, Horwitz, Horwitz 2007, Devine 2002). For this purpose, we quantitatively summarize the last twenty years of empirical research on diversity in organizational work teams (from 1994 to 2014). Based on a thorough literature review, we identify 242 empirical studies on the link between diversity and performance in organizational teamwork settings. Following the tradition of previous meta-analyses on workgroup diversity (e.g. Bell 2007, Bell et al. 2011, De Dreu, Weingart 2003, Horwitz, Horwitz 2007, Stewart 2006, Webber, Donahue 2001), we apply strict inclusion criteria to manage the vast amount of literature. Eventually, we include a total of 66 individual samples from 63 studies into the meta-analysis.

Overall, our findings confirm the ambivalent nature of diversity in teams. We identified the impact of diversity on team performance to depend primarily on the diversity attribute being either highly or less task-related. On the one hand, we observe a negative relationship between team performance and heterogeneity along less task-related attributes (gender and age). These findings largely support social category theory and similarity-attraction paradigm. On the other hand, we observe heterogeneity along highly task-related attributes (educational background) to be positively linked to team performance. This finding largely supports economic assumptions and information processing perspective. Furthermore, we observe both team size and team type to moderate the diversity-performance relationship. With increasing team size the influence of diversity on performance decreases for less task-related attributes and increases for highly task-related attributes. Concerning team type, we observe a positive moderating effect of top management teams (TMT) as well as research and development (R&D) teams on almost all diversity attributes irrespective of the task-relatedness. In contrast to this, work teams negatively moderate all diversity-performance relations. All results will be discussed and interpreted in detail throughout this study.

This paper proceeds as follows. Section two will discuss the diversity construct, its dimensions, the most common heterogeneity measures as well as a definition of diversity as understood in this paper. Subsequently, section three will present the two competing theoretical concepts and derive the main research hypotheses of this paper. The meta-analytic procedure applied in this study will be described in section four. Eventually, section five presents the results of the meta-analysis while section six concludes and offers implications for practical use and future research.

2.2 Diversity

Despite the relevance of diversity in the recent academic discourse, there is a lack of both a common understanding and a unique definition of the underlying construct (e.g. Guzzo, Dickson 1996). This becomes even more apparent considering the high number of synonyms that are used interchangeable for the umbrella term ‘diversity’ by scholars: “*’dispersion’, ‘inequality’, ‘within-group variability’, ‘(dis)agreement’, ‘consensus’, ‘heterogeneity’, ‘homogeneity’, ‘deviation’, ‘difference’, ‘distance’, ‘relational demography’,* *'sharedness' and more*" (Harrison, Sin 2006, p. 195). Since each separate definition is linked with a slightly different understanding and metric, the lack of a common diversity concept seems to be even more obstructive. Hence, any comparison across studies is complicated and needs particular attention due to the fact that researchers do not necessarily mean the same phenomenon when referring to diversity.

In general, three main inconsistencies have been identified in diversity research: the lack of a unique constitutive definition, the high variety within diversity variables and the inconsistent use of diversity measures (Harrison, Sin 2006). This paragraph addresses these inconsistencies, first by giving a clear definition of diversity as understood throughout this paper, and second by discussing those measures that are most commonly used to assess work group heterogeneity.

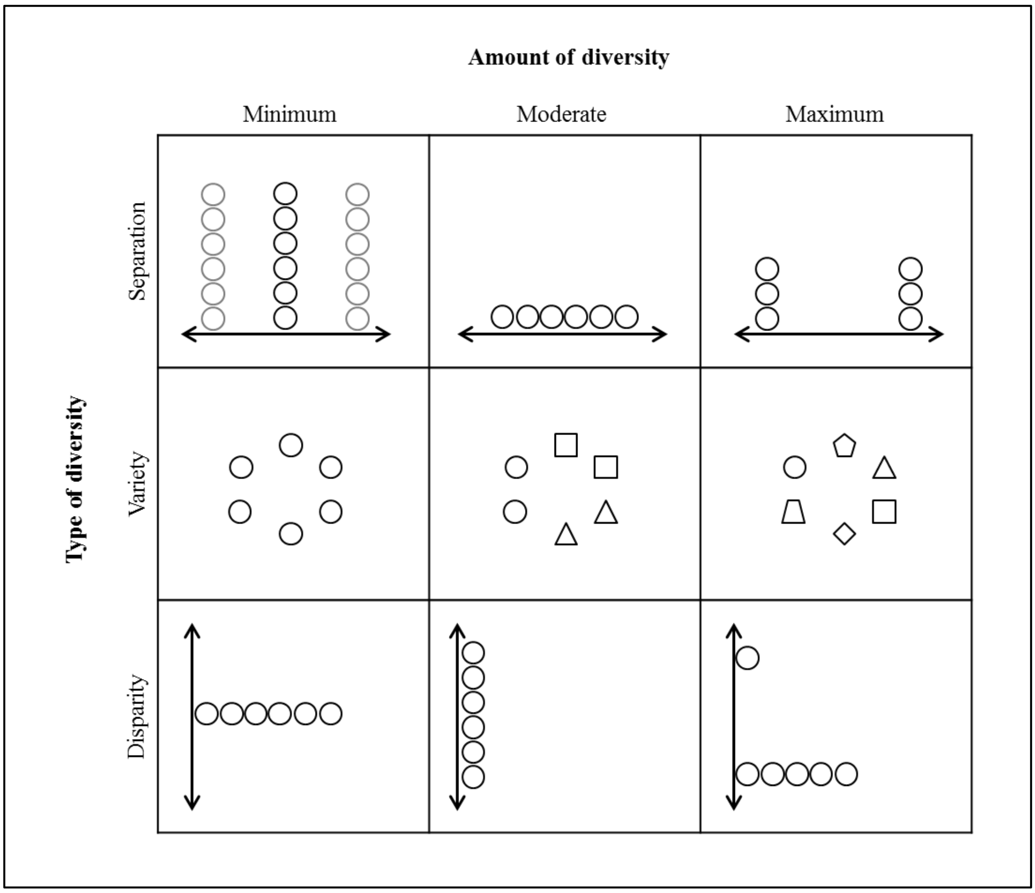
2.2.1 Definitions of Diversity

In many cases, studies on diversity focus exclusively on how diversity affects processes and outcomes of the units under observation. However, a controversial examination of the diversity construct itself is seldom found in the literature. Thus, it comes as no surprise that there is no joint agreement on how diversity can be defined. Instead, literature provides a plethora of divergent attempts that somehow dilute the understanding of diversity (Harrison, Sin, 2006).

A pioneering definition of diversity in its psychological, sociological and economic sense has been proposed in a seminal work by Blau (1977). According to his understanding, heterogeneity on a given attribute depends on its distribution: "*the larger the number of groups and the more evenly distributed the population is divided among them, the greater is heterogeneity*" (Blau, 1977, p. 9). However, this definition requires variables to allow for rank-ordering and, thus, is strictly associated with status or power hierarchy. Scholars ever since have broadened this narrow definition by providing more extensive diversity concepts (Harrison, Sin, 2006). The fundamental question remains the same: how is maximum diversity defined. While the idea of minimum diversity simply reflects perfect homogeneity of a group in terms of a given attribute, maximum diversity seems to have a two-folded dimension. On the one hand, let us assume a group with all members having different characteristics along a particular attribute. This group may be assessed to reach the maximum degree of diversity. On the other hand, consider a group that can be divided into two subgroups that are clearly separated along a few but extremely opposing characteristics. This group may be named maximal diverse as well. Moreover, one even might assume maximum diversity to appear in groups with one member clearly surpassing the others along a given characteristic. However, diversity literature has not put much attention on the clarification of these questions so far (Harrison, Klein 2007).

In this meta-analysis, we understand diversity according to a definition recently proposed by Harrison and Klein (2007). Addressing the ambiguity of existing ideas on diversity, they distinguish between three types of group heterogeneity: separation, variety and disparity (see Figure 1.1). First, diversity as separation expresses horizontal differences in position within a group along attributes such as attitudes, values or opinions. Second, diversity as variety describes divergent categories among team members on a given attribute such as information or educational background. Third, diversity as disparity assesses vertical differences within a group on valued social assets such as pay or hierarchy. In general, the authors regard diversity as “*the distribution of differences among the members of a unit with respect to a common attribute, X*” (Harrison, Klein 2007, p. 1200).

Figure 2.1: Diversity as Separation, Variety and Disparity



Source: Harrison, Klein (2007), p. 1202.

2.2.2 Dimensions of Diversity

Irrespective of its type, diversity is assumed to be attribute-specific. Put differently, a team is not diverse per se, but only regarding one or more clearly specified team member characteristics (Harrison, Klein, 2007). Any human being is a unique combination of features such as age, gender, cultural background, education, organizational tenure etc. In response to the large number of potential diversity attributes, scholars typically apply clustering in order to increase structure and manageability of data. The most fundamental cluster may be a distinction between easily observable attributes or surface level diversity (age, sex, and race) and less visible attributes or deep level diversity (tenure, education and function) (e.g. Harrison, Price, Bell 1998, Milliken, Martins 1996).

Another commonly used sorting of diversity attributes has been used by Polzer, Milton and Swann (2002) who distinguish between demographic (age, sex, race, citizenship) and functional features (previous degree, previous job function, MBA concentration). A slightly different logic is proposed by Pelled (1996) as well as Pelled, Eisenhardt and Xin (1999): They classify attributes to be either task-related (education, organizational tenure, group tenure, functional background) or non-task related (age, gender, race). In this study, we follow their recommendation by differentiating between less task-related and highly task-related diversity attributes.

2.2.3 Measures of Diversity

Usually, team diversity is assessed based on standardized heterogeneity measures at the group-level. The four most frequently used measures include the simple within-group standard deviation (s.d.), Blau's (1977) index of heterogeneity, Allison's (1978) coefficient of variation (CV) and Teachman's (1980) index of heterogeneity.

Within-group s.d. is one of the basic statistical measures to assess within-group variation in the distribution of a given attribute. It reaches its maximum in distributions that are bi-modal, i.e. the sample split in half on a given attribute (Harrison, Sin 2006). The Blau (1977) index is one of the most widely used measures to assess diversity of categorical variables such as gender or age (e.g. Timmerman 2000). Maximum heterogeneity is reached when the proportions of all possible categories are equally distributed. Therefore, the Blau index minimizes with total homogeneity and approaches one with increasing heterogeneity (Harrison, Sin 2006). Teachman's (1980) index may range from zero to positive infinity depending on the number of categories of any (categorical) variable. It is equal to zero in situations where all team members belong to the same social category and increases with the number of different categories in the team (Harrison, Sin 2006). Allison's CV (Allison 1978) is probably the most commonly used measure to assess within-group diversity (e.g. Williams, O'Reilly 1998). It is calculated by dividing the within-group standard deviation by the group mean. In contrast to other diversity measures, CV does not maximize when the variety of categories in the group increases but, likewise to s.d., grows with the magnitude of contrasts in the distribution of a given attribute. It reaches maximum values in teams with only one member being different to the others (Harrison, Sin 2006). The formulas for calculating these heterogeneity measures are presented in Table 2.1. However, all measures share a joint conceptual weakness as they are all prone to small team bias (Bedeian, Mossholder 2000) as well as distortions arising from divergent team sizes in only one sample (Biemann, Kerney 2010).

Table 2.1: Formulas for the Calculation of Heterogeneity Measures

Heterogeneity measure	**Formula**
Allison's (1978) Coefficient of Variation	$CV = \frac{\sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N}}}{\bar{x}}$
Blau's (1977) Index	$B = 1 - \sum_{k=1}^{N} p_k^2$
Standard Deviation	$SD = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N}}$
Teachman's (1980) Index	$T = -(\sum_{k=1}^{N} p_k \ln(p_k))$

Note. N = total number of teams; x_i = individual level characteristics; $\bar{x}$ = mean characteristics of the team; p_k = proportion of team members in the *k*th category.

Apart from the group-level heterogeneity measures presented above, there are other ways to conceptualize heterogeneity. First, relational diversity describes an individual's (dis)similarity with his team mates and, thus, measures diversity at the individual level (e.g. O'Reilly, Caldwell, Barnett 1989, Tsui, Egan, O'Reilly 1992, Harrison, Price, Bell

1998). Second, research using within-group faultlines to assess group heterogeneity is recently increasing. The faultlines approach identifies potential subgroups in teams and evaluates similarity within as well as dissimilarity between subgroups (e.g. Lau, Murnighan 1998, Shaw 2004, Thatcher, Pantel 2011). However, in this meta-analysis we exclusively focus on group-level heterogeneity and do not include studies that apply either relational diversity or the faultlines approach.

2.3 Theoretical Background of Diversity Research

Diversity in teams is often referred to as "*double-edged sword*" (Milliken, Martins 1996, p. 403) since two opposing research traditions and conflicting empirical findings do not allow for universally valid conclusions on the effects of homogeneity and heterogeneity in teams. On the one hand, social categorization theory and similarity-attraction paradigm emphasize the destructive nature of heterogeneity to team processes and performance. On the other hand, economic theory and information processing perspective argue that heterogeneity may be beneficial to team performance.[6] Both theoretical strands are backed up by broad empirical and experimental evidence. The upcoming sections will briefly introduce the underlying ideas and main contributions.

2.3.1 Social Categorization and Similarity-Attraction Paradigm

Social categorization is the most commonly used theory in demography research. It emphasizes the negative consequences of heterogeneity in teams on team processes, outcomes and performance (Williams, O'Reilly 1998). The social categorization concept was established in the early 1980s and has mainly been shaped by Turner (1987). It extends the social identity concept developed some years earlier by Tajfel (1982). While social identity focuses on intergroup behavior and seeks to explain the psychological processes of intergroup discrimination, the social category theory is centered more generally on the question of how people may function as a group at all (Turner 1987). In that, it includes the social identity concept. As both theories share the same logic, this review only applies the more commonly used term of self-categorization for simplification.

[6] In addition, there are contributions arguing that both theoretical perspectives do not necessarily contradict each other but may simultaneously account for positive and negative effects of diversity in workgroups. One approach is the categorization-elaboration model (CEM) that incorporates assumptions of both information processing and social categorization (e.g. van Knippenberg, De Dreu, Homan 2004, van Knippenberg and Shippers 2007).

The main idea is simple: Self-categorization is based on the assumption that individuals define their social identity through their affiliation to social groups that are associated with emotional significance to them (Tajfel 1972, 1982). According to McGrath, Berdahl and Arrow (1995), it is this understanding that is colloquially referred to as diversity by most people. Usually, individuals have a strong desire to maintain a positive self-identity as well as a high level of self-esteem (Tajfel, Turner 1986). Both self-identity and self-esteem typically occur through social comparison with peers. Human beings, therefore, subconsciously sort themselves and others into social categories along any possible dimension (e.g. age, gender, race, nationality, religion, organizational membership etc.). In that sense, they define their individual self-identity in comparison to others (Tajfel, Turner 1986).[7] Based on this process, subgroups – in-groups vs. out-groups – emerge along the underlying characteristics (Tajfel 1969). This implies a process of depersonalization as people are no longer seen as individuals, but as members of a given category (Brown, Turner 1981). This in-group-out-group distinction manipulates the individual perception of behavior (Guinote, Fiske 2003) in favor of the in-group (Kulik, Bainbridge 2006, Bilig, Tajfel 1973). Thus, any individual stereotype dissociates from members of other categories (Tajfel 1982). As a consequence, out-group members are – even when classified on arbitrary variables – seen as less trustworthy, honest and cooperative (Brewer 1979). This so-called in-group bias was confirmed by Tajfel (1982), who detects this bias in reviewing over thirty studies. However, as people mirror negative behavior of in-group members (e.g. stereotyping or dissociation), hostile interactions as well as problematic inter-subgroup relations may occur and lead to an “us vs. them”- mentality within the group (van Knippenberg, de Dreu, Homan 2004). As a consequence, members that are similar along demographic traits are more likely to cooperate and communicate with each other. As a result, homogeneous groups should systematically outperform heterogeneous teams. Hence, social categorization is mainly referred to by diversity research that emphasizes team heterogeneity to be destructive to team performance and outcome (see Horwitz, Horwitz 2007 for a review of studies).

[7] Turner (1987) distinguishes between three levels of abstraction within the social self-categorization concept. First, individuals categorize themselves as human beings; this is referred to as super-ordinate level. Second, individuals categorize themselves into social groups along the distinction of in-group vs. out-group; this is referred to as intermediate level. Third, individuals differentiate between their self as a unique person in comparison to the other in-group members; this is referred to as subordinate level. The levels are also referred to as human (inter-species), social (inter-group) and personal identity (interpersonal comparison) levels, respectively.

Similar to social categorization, the similarity-attraction paradigm – introduced by Byrne (1969, 1971) – predicts homogeneous teams to outperform heterogeneous teams. The paradigm is built on the assumption that people tend to be interpersonally attracted to others that are – or are perceived to be – similar to them along demographic traits, activities, attitudes or values (Byrne 1969, 1971). Vice versa, dissimilarity decreases interpersonal attraction (Rosenbaum 1986). The similarity-attraction concept differs from social categorization in that it focuses entirely on the attraction of similarity and does not include any negative or hostile stances towards non-similar others, e.g. out-group members (Ely 2004). In work settings, similarity facilitates cooperation among interpersonally attracted individuals and, hence, may function as positive reinforcer (Williams, O'Reilly 1998). Moreover, similarity and high interpersonal attraction help people to simplify their communication that in turn leads to high social integration (Wiersema, Bantel 1992, Tsui et al 1992). The logic of similarity-attraction is well established in the literature as it has been demonstrated that individuals are more likely to select others that are similar to themselves in free choice situations and that, as a consequence, naturally formed groups in organizations are typically based on similarity (e.g. Ancona, Caldwell 1992). Consequentially, the similarity-attraction effect is acknowledged to be "*one of the most robust findings in social psychology*" (Martins et al. 2003, p. 78).

Given the high visibility and salience of demographic attributes such as age, gender and cultural background, these less task-related surface-level attributes can be assumed to provoke social categorization and similarity-attraction more easily than other, deep-level demographic attributes. As a result, heterogeneity in age, gender and cultural background may hamper team cooperation and decrease team performance. A substantial body of experimental and empirical literature confirms these arguments (e.g. Brewer 1979, O'Reilly, Caldwell, Barnett 1989, Chatman, Flynn 2001, Wiersema, Bantel 1992, Williams, O'Reilly 1998).[8] Therefore, based on social categorization and similarity-attraction paradigm, the following hypotheses can be derived with regard to less task-related demographic attributes:

[8] Yet, the impact of divergent surface-level demographic attributes is dynamic and may change with the length of cooperation due to two main effects. First, negative categorization effects may deteriorate as other non-salient categories surface after some time (Harrison, Price, Bell 1998). Second, team members start to classify themselves as fellow in-group members and increase cooperation (Chatman, Flynn 2001).

H1a: Age diversity is related negatively to overall team performance.

H1b: Gender diversity is related negatively to overall team performance.

H1c: Cultural background diversity is related negatively to overall team performance.

2.3.2 Information Processing Perspective and Economic Theory

In contrast to the aforementioned concepts that emphasize the superiority of homogeneous teams, the information processing perspective perceives a value in diversity. Following this idea, team members may benefit from combining their complementary individual resources in order to balance individual weaknesses within the team. As a result, heterogeneous teams are supposed to outperform homogenous groups since members of the latter are limited to substituting resources (Cox, Lobel, McLeod 1991, Nemeth 1986, Stasser, Stewart, Wittenbaum 1995).

The underlying theoretical argument was first introduced by Hoffman (1959) and Hoffman and Maier (1961) and advanced among others by Cox and Blake (1991) as well as Cox (1993). They all build on the idea that diversity within groups along a given set of attributes such as cognitive resources, functional background, experiences, tenure, information, education and network ties will positively influence group performance – at least in tasks associated with cognitive ability. Thus, the value of heterogeneity in groups derives from a broad base of resources such as knowledge and information as well as a more diversified set of methods to face a task – both not available to homogeneous groups (Cox, Blake 1991, Hambrick, Cho, Chen 1996, van Knippenberg, Schippers 2007). Furthermore, varying perspectives, experiences or viewpoints – especially when largely contrasting (Nemeth 1986) – can stimulate vivid discussions within the group and, thus, lead to new insights, ideas and solutions (Argote, Gruenfeld, Naquin 2001, Levine, Resnick, Higgins 1993). Additionally, members of heterogeneous teams have complementary network ties with people outside the initial group who provide external perspectives and additional information from various sources (Ancona, Caldwell 1992, Jehn, Nothcraft, Neale 1997, Zenger, Lawrence 1989). For instance, one can easily observe an inverted cohort effect with regard to heterogeneity in organizational tenure since members that have entered the organization at different points in time have established diversified information networks, communication patterns, skills and experiences made within the organization (Ancona, Caldwell 1992). As a consequence, teams that are heterogeneous along these attributes are often found to be more effective decision-makers, better problem solvers, superior product designers and more creative innovators (e.g. Ancona, Caldwell 1992, Pelled, Eisenhardt, Xin 1999, Wanous, Youtz 1986).

Lazear (1999) offers an economic approach modeled on three factors that determine productivity gains from heterogeneous teams. First, a team can only benefit from heterogeneous team members if the individual skills or information sets are not overlapping but disjoint in the way that cooperation extends the information or skills available to the individual. Exemplarily, Lazear (1999) supposes two economists who have a common educational and professional background and, therefore, might not stimulate each other's work since the information they may share is largely substitutive. As soon as heterogeneous people come into play complementary information becomes available leading to positive effects on team performance, in particular if tasks within the team are complementary (Prat 2002). Second, the individual skills or information sets must be relevant to the other team members and to the team task respectively. Lazear (1999) thinks of an auto mechanic whose knowledge is – although being disjoint from an economist's knowledge – not relevant to the economist's work and, therefore, does not offer any help. In this case, diversity does not yield any economic gain. Third, diverse skills or information sets are only helpful if group members are able to communicate easily with each other to exchange the relevant information at a low cost. Without a common language, exchange would be costly and may wipe out possible gains coming from disjoint and relevant skills / information. As a consequence of high communication costs, people tend to communicate less frequently, which is negatively associated with firm performance, especially in businesses where cooperation is necessary – e.g. in creative tasks (Zenger, Lawrence 1989). Although Lazear (1999) limits his understanding of language to its literal sense (e.g. English or French), other authors have extended the idea to other features of communication as well. For instance, Kearney, Gebert and Voelpel (2009) state that professional jargons or varying interpretative schemata may also impede communication.

Based on the information processing perspective as well as Lazear's (1999) economic considerations, we assume highly task-related diversity to have a strong positive effect on team performance. This leads to the following hypotheses:

H1d: Tenure diversity is related positively to overall team performance.

H1e: Functional background diversity is related positively to overall team performance.

H1f: Educational background diversity is related positively to overall team performance.

H1g: Education level diversity is related positively to overall team performance.

2.3.3 Moderators of the Diversity-Performance Relationship

Diversity research – even if limited to primary studies on diversity at the work group-level – is still heterogeneous with regard to team size and team type. Both team size and team type might affect the diversity-performance relationship. We, therefore, hypothesize potential moderating effects.

Team size

It is a stylized fact that large teams have a higher probability of being heterogeneous than smaller teams. Thus, one might conclude that team size affects the diversity-performance relationship since team dissimilarity increases with additional team members. Indeed, recent research identifies team size to be a significant moderator of the relationship between team diversity and team outcomes (e.g. Bowers, Pharmer, Salas 2000, Stewart 2006, Wegge et al. 2008). Yet, other studies fail to confirm the moderating effect of team size in the diversity-outcome relationship (e.g. Horwitz, Horwitz 2007, Wiersema, Bantel 1992). This dichotomy of findings may be attributable to two opposing ways of how team size may affect team performance. Each additional team member may contribute additional skills and knowledge that may complement those already available in the team. This broadens the cognitive resources of a team and, therefore, can be assumed to be beneficial to team performance (e.g. Hambrick, Cho, Chen 1996, van Knippenberg, Schippers 2007). Nevertheless, the risk of negative dissimilarities that are observed to be detrimental to communication, group cohesion, coordination and cooperation among team members may increase with team size (e.g. Amason, Sapienza 1997, Chatman, Flynn 2001, Smith et al. 1994). As a result, team performance decreases.

In this study, we assume a positive moderating effect of team size on the diversity-performance relationship to be valid for highly task-related demographic characteristics such as tenure, function and education. Likewise, we believe the less task-related characteristics such as age, gender or cultural background to negatively moderate the link between group diversity and performance. Findings by Stewart (2006) support this argument. He confirms the positive moderating effect for top-management teams and project teams while observing a negative effect for production teams. Being aware of the cognitive nature of team processes in TMT and project teams in contrast to the repetitive and standardized work of production teams, these results come as no surprise. Based on these considerations, we predict:

H2a: Team size negatively moderates the relationship between diversity and team performance for less task-related demographics.

H2b: Team size positively moderates the relationship between diversity and team performance for highly task-related demographics.

As discussed above, team size might moderate the diversity-performance relationship simply for statistical reasons since the most commonly used heterogeneity measures (e.g. Blau index of heterogeneity, coefficient of variation, Teachman index) are prone to biases arising either from small team size (Bedeian, Mossholder 2000) or from team size varying within the sample (Biemann, Kerney 2010). The necessity to test for statistical artefacts in the diversity-performance relationship affirms us in assessing team size as a potential moderator of the diversity performance relationship.

Team type

It is a well-established fact in the team literature that team performance and the efficiency of teamwork is highly dependent on the type of the team (e.g. Cohen, Bailey 1997, Thylefors, Persson, Hellström 2005). At the same time, there is broad evidence that team type moderates the diversity-performance relationship (e.g. Horwitz 2005, Stewart 2006). Team research applies multiple ways to distinguish between team types. For instance, Joshi and Roh (2009) differentiate teams along the longevity of cooperation. They assume visible diversity (e.g. age, gender) to be of less significance in teams designed for long-term cooperation (such as TMT or work groups deeply established in the organization), while invisible diversity (e.g. value, personality) is positively linked to length of cooperation (Harrison et al. 2002). Another strand of literature distinguishes teams along the tasks they perform (Sundstrom et al. 2000).

In this study, we follow a team typology proposed by Devine (2002). First, intellectual work teams (such as TMTs) are assigned complex tasks in often uncertain situations and are running non-standardized processes. Second, design teams (such as R&D) require creativity and technical innovation to fulfill their tasks. Third, physical work teams (such as production teams, performance teams or service teams) simply follow standardized tasks. We aim at testing for moderating effects along these three divergent team types. In general, TMTs and R&D teams are assumed to be heterogeneous on task-related attributes (i.e. functional and educational background) as the respective tasks require a broad base of varying knowledge and experience. At the same time, these teams are likely to be homogeneous on non-task related characteristics (i.e. age, gender, cultural background). In contrast to this, physical work teams are found to be more heterogeneous in their composition (e.g. Horwitz, Horwitz 2007, Devine 2002). As a consequence of these considerations, we derive two hypotheses on the impact of less task-related as well as highly task-related heterogeneity on team performance:

H3a: The predicted negative relationship between
less task-related heterogeneity and team performance will be weaker
in R&D teams and TMT compared to other team types.

H3b: The predicted positive relationship between
highly task-related heterogeneity and team performance will be stronger
in R&D teams and TMT compared to other team types.

2.4 Meta-Analysis Approach

In the following, we use the meta-analysis approach to test for the above stated hypotheses on the diversity-performance relationship as well as to examine potential moderating effects of team size and team type.

Literature review

A thorough literature review was conducted in order to identify studies on workgroup diversity. We focus on research published between 1994 and 2014 since we aim at quantitatively summarizing the last 20 years of diversity research without replicating earlier meta-analyses that have already comprehensively reviewed earlier works (e.g. Bowers, Pharmer, Salas 2000, Webber, Donahue 2001). We located relevant studies on workgroup diversity for potential inclusion through several search strategies. First, we reviewed recent meta-analyses and narrative reviews by Bower, Pharmer, Salas (2000), Williams, O'Reilly (1998), Horwitz, Horwitz (2007), Stewart (2006), Webber, Donahue (2001), Bell (2007), Bell et al. (2011), De Dreu, Weingart (2003) and De Church, Mesmer-Magnus (2010).

Second, we electronically searched the online databases *Scopus*, *Business Source Complete* and *PsycINFO* with combinations of the following keywords: *team, group, diversity, heterogeneity, homogeneity, diverse, (team) composition, demography, demographic diversity, age, gender, female, male, tenure, education, function, culture, race, ethnicity, performance, outcome, effectiveness and top-management-team.* Third, the database search was supplemented with a manual search for in-press articles in both top-tiered journals and journals with a particular focus on group research including *Administrative Science Quarterly*, *Organization Science*, *Journal of Applied Psychology*, *Academy of Management Journal*, *Academy of Management Review*, *Personnel Psychology*, *Group and Organizational Studies/Management*, *Organizational Behavior and Human Decision Processes*, *Journal of Organizational Behavior*, *Journal of Management*, *Small Group Research*, *Group Dynamics* and *Human Relations*. We only consider peer-reviewed, English-speaking journal articles, thus excluding dissertations, conference papers, manuscripts or working papers. We are aware of the risk that this approach might put our results into question due to information loss as a consequence of publication bias. Yet, we proceed as described to be able to deal with the garbage-in-garbage-out issue (e.g. Egger, Smith, Sterne 2001) that seems of particular importance to us in the wide and complex field of diversity research. However, since published findings feature positive as well as negative relationships and do also report non-significant results at some instances, we feel confident to catch the substantial contributions without losing relevant information. Study abstracts were reviewed for suitable research on the relationship between demographic diversity variables and both team performance and creativity. This literature search yields a total of 242 studies that were further examined for potential inclusion in the meta-analysis.

In order to be included, a study has to meet several inclusion criteria. First, we focus exclusively on empirical research that studies at least one of the hypothesized relationships between demographic diversity variables and workgroup productivity. In other words, we exclude studies that observe outcomes such as creativity, innovativeness, satisfaction or group cohesion. Additionally, we refrain from including aggregated measures of diversity such as demographic diversity, which is often used to combine age, gender and cultural background/ethnicity in only one variable. Moreover, we do not consider studies that observe psychological diversity variables such as cognitive ability or personality. Second, we only include studies that understood work groups as teams that meet the following criteria: Team members need to share a common goal, have similar or even identical working tasks and be working interdependently. Moreover, teams should be embedded in an overall social system such as an organization and be regarded as a team by its members as well as by external persons (Guzzo, Dickson 1996). Third, since mixed levels of analysis are inappropriate for calculating sample-weighted effects (Beal et al. 2003), we focus on studies which assess diversity on the team-level and drop those studies that report findings on the individual level. Furthermore, we follow suggestions by Katzenbach and Smith (2005) and limit the maximum team size to twenty-five persons. Members of larger teams are usually not subject to task interdependency and do not adequately interact as a team, yet both factors are necessary to study diversity effects on team performance. Minimum team size is two persons. Fourth, we consider only studies that assess diversity based on objective measures such as within-group s.d., Blau's (1977) index of heterogeneity, Allison's (1978) CV and Teachman's (1980) index of heterogeneity. Subjective measures of group heterogeneity are most likely biased due to individually differing perceptions of current group settings or cultural variations and are, therefore, not included in this meta-analysis. Additionally, we exclude studies that understand diversity either as relational diversity or as faultlines since both approaches are not comparable to the diversity concept applied here. Fifth, studies have to stem from real-life organizational settings. This excludes experimental research, case studies and simulation research design. Moreover, we concentrate on teams in working environments. Therefore, we exclude studies that are based on student work groups, sport teams or other non-task related organizational settings. We do so in order to guarantee high practical relevance of implications drawn from this meta-analysis. Eventually, we exclude studies that we asses to suffer methodological weaknesses and/or are not comparable to the overall set of studies. Applying these strict inclusion criteria is necessary to guarantee between-study comparability. A total of 179 studies out of 242 initially identified studies failed to meet the inclusion criteria. This leaves a final number of 63 studies to be included in our analysis. Two studies contribute two or more samples. As a result, we yield a final data set including 66 samples with information on a total of 12,478 teams.

Diversity variables

The independent variables investigated in this meta-analysis include the demographic features of age, gender, cultural/ethnical background, organizational tenure, functional and educational background as well as education level. To meet the inclusion criteria, diversity variables have to be aggregated to the team-level by means of an appropriate measure of group-level heterogeneity, e.g. within-group s.d., the standard measures Blau's (1977) index of heterogeneity, Allison's (1978) CV and Teachman's (1980) index of heterogeneity. Still, we allow for other methods to assess diversity objectively – such as the Herfindal index, the Gini index or the simple percentage share – as long as its choice seems reasonable and comparability with other measures is given. Table A.1 in the appendix lists the frequency distributions of diversity measures used in primary studies.

Following Pelled (1996) in clustering diversity attributes along the dimension of task-relatedness, we categorize age, gender and cultural background/ethnicity (referred to as culture in the following) as less task-related, while we rate organizational tenure, functional and educational background as well as education level to be highly task-related. While the metric of age, gender, culture and organizational tenure diversity is quite self-explanatory and only varies slightly between primary studies (e.g. age and tenure measured continuously or in categories), three diversity variables need further consideration. Functional diversity is understood as functional experience gained during the career, the dominant function, organizational roles and the primary professional orientation. Educational background diversity is conceptualized as the major level or the specialization of studies. Education level diversity is assessed either as the number of years of (formal) education or as the highest educational level achieved.

Outcome variables

The dependent outcome variables include divergent types of team performance assessed along either objective or subjective performance measures. Objective performance measures include financial key figures such as return on equity (ROE), return on assets (ROA) and return on sales (ROS) as well as measures of team effectiveness such as productivity, profitability and goal achievement. Subjective performance measures are based either on a team's self-assessed performance appraisals or on external ratings from managers, supervisors or other persons outside the teams.

Moderator variables

Eventually, in order to test for potential moderating effects in the diversity-performance relationship we include information on average team size and team type. We include mean team size when reported in the initial study – usually measured as simple headcount of team members. Studies that do not report average team size are excluded from the moderator analysis. Regarding team type, we cluster teams along Devine's (2002) typology either to be TMT, work team, R&D team or mixed teams. While TMT (executive) and R&D teams (design) match Devine's (2002) intellectual work team cluster, work teams refer to the physical work team cluster. We understand "work teams" to include those teams introduced as work teams in the initial studies, but also teams that continuously work together on the same persisting tasks, e.g. teams labeled as school staff, sales teams or branch teams. Studies on work teams that are engaged in specific types of cooperation (e.g. virtual teams) or subject to task profiles that change over time (e.g. venture or project teams) are classified as "mixed teams". The same is true for studies which include divergent team types without reporting results separately. Again, studies that do not provide the appropriate information on team type are excluded from moderator analysis at this instance.

Meta-analytic procedure

For this meta-analysis, we chose the product-moment correlation coefficient (r) to be the primary effect size index, because we exclusively focus on observational research and do not include experimental findings. However, product-moment correlation coefficients may entail serious statistical difficulties such as problematic standard error formulation (Lipsey, Wilson 2001) and excessive Type I error rates (Alexander, Scozzaro, Brodkin 1989). In response to these undesirable statistical properties, we follow suggestions by Lipsey and Wilson (2001) as well as Hedges and Olkin (1985) and correct correlation coefficients by Fisher's Z_r-transformation.

With regard to model specification, we chose random effect models as introduced by Hedges and Olkin (1985) since this approach allows heterogeneity in the effect size to arise from two sources, subject level sampling error and variability of effects randomly distributed along studies (Lipsey, Wilson 2001). To test for the degree of precision of the mean effect size estimates, we calculated the 95% confidence interval (CI) around the population correlation. A CI that does not include zero indicates the mean effect size to be significant (Lipsey, Wilson 2001, Whitener 1990). If the 95% CI does not include zero, we follow Lipsey and Wilson (2001) in calculating *z*-tests to check whether the mean population effect size is significant at $p \leq \alpha$ (e.g. 10%-significance). In order to test for the homogeneity of the effect size, we apply chi-square distributed Q statistics (Hedges, Olkin 1985). In general, homogeneity of effect sizes would imply that individual effect sizes differ from the estimated population mean effect size only by sampling error (Lipsey, Wilson

2001).[9] In other words, rejecting Q statistics would denote heterogeneous effect size distributions with two main implications. First, heterogeneity favors the use of random effects rather than fixed effects. Since we observe the majority of Q statistics to confirm heterogeneity in our data, our initial choice of random effect models is statistically supported. Second, heterogeneity suggests that randomly distributed variables may moderate the variance in individual effect sizes. As a consequence, we test for the proposed moderating effects of team size and team type by applying the DerSimonian and Laird (1986) random-effects meta-regression methods approach with Knapp and Hartung (2003) tests of effect estimates variance. Q_E is the corresponding residual heterogeneity statistic that indicates model significance (Lipsey, Wilson 2001).

2.5 Results

We start the discussion of our findings by examining the relationship between diversity variables and team performance as hypothesized in H1a to H1g (Table 2.2). Subsequently, we present our findings with respect to the moderating effects of average team size (H2a, H2b, Table 2.3) as well as team type (H3a, H3b, Table 2.4). For each analysis, we report the number of correlations from independent samples included in the meta-analysis (*k*), the total number of teams (*N*), the effect size given as the corrected population correlation (r_z) (sample-size weighted through Fisher's *Z*-transformation), the standard error of the corrected population correlation ($r_z(se)$), the lower and upper bounds of the 95% confidence interval (*95% CI*), Cochran's Q as homogeneity statistics (*Q*) with the respective degree of freedom (*df*), and eventually the I^2 statistic which presents the percentage of between-study variation attributable to heterogeneity (I^2).

Main results on the diversity-performance relationships (hypotheses H1a to H1g) are reported in Table 2.2. Forest plots that illustrate the correlations a study contributes to the meta-analysis are presented for each diversity variable in Figures A.1 to A.7 in the appendix. In support of Hypothesis H1a, we observe a negative effect size regarding the link between age heterogeneity and team outcome (r_z = -0.052, 95% CI: -0.109 | 0.005). Since zero is included in the 95% CI, we calculate the *z*-test of effect size (ES) which allows to

---

9 As an alternative to Q-statistics we calculate the I^2 measure that gives the proportion of total variation in mean effect size attributable to between-study heterogeneity (Higgins, Thompson 2002). It is calculated as $I^2 = 100\%\ x\ \frac{Q-df}{Q}$. As rule of thumb, one can say that an I^2 value around 25% indicates low heterogeneity, while values around 50% and 75% can be translated to moderate and high levels of heterogeneity, respectively (Higgins et al. 2003).

reject the null-hypothesis (ES=0) and, thus, confirm H1a at the 10%-level of significance. In other words, the more diverse team members are with regard to age the lower is overall group productivity. Hypothesis H1b predicted a negative relationship between gender heterogeneity and team performance. Our findings confirm the expected link (r_z =-0.036, 95% CI: -0.076 | 0.004). Again, *z*-tests of effect size (ES) allow for confirmation of H1b (at the 10%-level). This indicates that gender heterogeneity in teams is detrimental to team performance. Regarding the relation between cultural diversity and team performance, we do not find any support for the expected negative relationship (H1c). Cultural diversity does not affect team productivity significantly. In summarizing findings on the less task-related diversity variables, results for both age and gender provide support for social categorization theory as well as the similarity-attraction paradigm since diversity along these attributes is identified to affect overall team performance negatively.

Table 2.2: Relationship between Diversity and Overall Team Performance

Diversity	k	r_z	r_z(se)	95% CI Lower	95% CI Upper	Q	df	I^2
Age	25	-0.052*	0.029	-0.109	0.005	88.79***	24	73.00%
Gender	30	-0.036*	0.021	-0.076	0.004	57.69***	29	49.70%
Culture	20	0.003	0.033	-0.061	0.067	105.98***	19	82.10%
Function	27	0.041	0.039	-0.035	0.117	369.37***	26	93.00%
Tenure	25	0.007	0.031	-0.054	0.069	178.85***	24	86.60%
Education								
Background	10	0.064*	0.038	-0.011	0.138	48.85***	9	81.60%
Level	12	-0.006	0.052	-0.108	0.097	43.23***	11	74.60%

Note. *k* = total number of correlation coefficients meta-analyzed; r_z = corrected population correlation (sample-size weighted based effect size (ES) on Fisher's z transformed correlation coefficients with significance test for ES=0; r_z = standard error of the corrected population correlation; *95% CI* = lower and upper bound of the 95% confidence interval; Q = homogeneity statistics; df = degree of freedom; I^2 = percentage of between-study variation due to heterogeneity.
***p<.01; **p<.05; *p<.1

With regard to the expected positive relationship of heterogeneity in highly task-related attributes on team productivity, we only find support for educational background diversity as stated in H1f (r_z = 0.064, 95% CI: -0.011 | 0.138). A significant z-test allows us to confirm H1f at the 10%-level of significance. This finding suggests teams to perform significantly better when drawing on a broad set of diverse educational backgrounds. However, there is no support for the predicted positive relationship between team performance and organizational tenure, functional and education level heterogeneity (H1d, H1e and H1g).

Therefore, overall support for the information processing perspective is weak and only based on findings with respect to educational background heterogeneity.

In order to investigate the relationship between team heterogeneity and team performance in more depth, we separate primary studies depending on the way they assess team performance (Table A.2 in the appendix). While there are no outcomes other than those discussed above in the fields of age, gender, cultural, functional and educational background diversity, we gain additional insights for tenure diversity as well as education level diversity. In partial support of Hypothesis H1d, we observe tenure diversity to be positively related to team performance at least when performance measures are based on subjective external ratings, e.g. by supervisors (r_z = 0.095, 95% CI: 0.021 | 0.106). Contrary to our expectations as expressed in hypothesis H1g, we found a negative relationship between educational level diversity and performance measured on outsider ratings (-0.132, 95% CI: -0.267 | 0.002).

All in all, we are able to confirm the expected negative link between team performance and both age diversity (H1a) and gender diversity (H1b). Furthermore, we find support for the hypothesized positive relationship between educational background diversity (H1f) and team performance. Additionally, we can provide partial support for the expected positive relation between team performance and tenure diversity (H1d). At the same time, we have to partially reject H1g as we observe a negative relationship between education level and performance rated by outsiders. Subsequent to the discussion of the main results of the meta-analyzed diversity-performance relationship, the second set of analyses centers on the potential moderating effects of average team size (as stated in H2a and H2b) and team type (as stated in H3a and H3b).

Team size

In order to test for the moderating effect of team size on the diversity-performance relationship, we applied random-effects meta-regression. Results are reported in Table 2.3. As stated in H2a, we expect team size to negatively moderate the link between heterogeneity in the less task-related demographic attributes of age, gender and culture. Meta-regression findings largely support team size to be the predicted negative moderator. In particular, the correlation coefficients for the diversity-performance relationship reduce with each additional unit member (for age by -0.001, for gender by -0.013 and for culture by -0.021). Highly significant values of Q_E confirm the significance of the regression models. These findings may suggest that coordination and communication among team members are complicated with increasing team size. Moreover, social categorization and conflict may be more pronounced in larger teams. As predicted in H2b, we observe positive moderating effects on the diversity-performance relationship for the task-related demographic attributes tenure (0.003), functional (0.013) and educational background (0.063) as well as education level (0.016). These findings suggest that additional team members may contribute additional (complementary) knowledge, skills and experience. This seems to be beneficial to group performance at least for task-related heterogeneity. Overall, we are able to confirm hypotheses H2a and H2b.

Table 2.3: Team Size as Moderator of the Performance-Diversity Relationship

Diversity	k	r_z	r_z(se)	95% CI Lower	95% CI Upper	Q_E	df	I^2
Age	19	-0.001	0.008	-0.019	0.017	62.31***	17	72.72%
Gender	21	-0.013	0.005	-0.024	-0.001	35.54**	19	46.54%
Culture	11	-0.021	0.013	-0.051	0.008	48.11***	9	81.29%
Function	15	0.013	0.016	-0.021	0.048	79.24***	13	83.59%
Tenure	14	0.003	0.022	-0.045	0.051	56.19***	12	78.65%
Education								
Background	6	0.063	0.049	-0.074	0.2	16.03***	4	75.05%
Level	11	0.016	0.026	-0.042	0.073	32.86***	9	72.61%

Note. *k* = total number of correlation coefficients meta-analyzed; r_z = corrected population correlation (sample-size weighted based ES on Fisher's z transformed correlation coefficients with significance test for ES=0; r_z = standard error of the corrected population correlation; *95% CI* = lower and upper bound of the 95% confidence interval; Q_E = homogeneity statistics; df = degree of freedom; I^2 = percentage of between-study variation due to heterogeneity.
***p<.01; **p<.05; *p<.1

Team type

Likewise to team size, we use meta-regression to test for moderating effects of team type on the diversity-performance relationship. Results are reported in Table 2.4. In hypothesis H3a we expected the negative relationship between less task-related heterogeneity and team performance to be weaker in TMT and R&D teams compared to other team types. For all three less task-related diversity attributes (age, gender and culture), we observe the predicted negative diversity-performance relationship for work teams and mixed teams. However, to our great surprise, results for TMT and R&D teams do not show the expected slightly weaker negative relationship but, instead, even suggest a positive relationship of age, gender and culture with team performance. In other words, for TMT and R&D teams the assumptions derived from social categorization as well as similarity-attraction paradigm on the advantageousness of homogeneous teams have to be neglected. Instead, our results support the information processing perspective stating that heterogeneity even in less task-related attributes is beneficial for TMT and R&D team performance. We attribute this finding to two possible explanations. First, as stated above, homogeneity in TMT and R&D teams along surface-level demographic characteristics is found to be very high (e.g. Horwitz, Horwitz 2007, Devine 2002). Second, we assume that within TMT and R&D teams disadvantages associated with heterogeneity (such as conflict, coordination as well as communication problems and social categorization) are of minor importance. Instead, even little heterogeneity in age, gender or culture seems to entail advantages to the particular tasks of TMT and R&D teams since strategic decision making and innovative thinking may be assumed to improve with differing perspectives of team members. For instance, consider management boards with and without female participation. In summary, we are able to confirm hypothesis H3a insofar as TMT and R&D teams not only suffer less from diversity in less task-related attributes but even benefit from the resulting divergent perspectives.

Table 2.4: Team Type as a Moderator of the Performance-Diversity Relationship

Team Type	k	r_z	r_z(se)	95% CI Lower	95% CI Upper	Q_E	df	I^2
Age								
TMT	20	0.023	0.064	-0.112	0.157	45.99***	18	60.86%
Work	20	-0.015	0.065	-0.152	0.123	44.44***	18	59.50%
R&D	20	0.185	0.15	-0.13	0.5	67.77***	18	73.44%
Mixed	20	-0.139	0.117	-0.386	0.107	68.64***	18	73.78%
Gender								
TMT	26	0.062	0.048	-0.036	0.1611	48.5***	24	50.52%
Work	26	-0.084	0.043	-0.173	0.0051	46.76***	24	48.67%
R&D	26	0.298	0.128	0.034	0.563	44.72***	24	46.33%
Mixed	26	-0.022	0.098	-0.225	0.181	51.29***	24	53.21%
Culture								
TMT	16	0.17	0.049	0.064	0.275	39.32***	14	64.40%
Work	16	-0.158	0.052	-0.269	-0.046	40.85***	14	65.73%
R&D	---	---	---	---	---	---	---	---
Mixed	16	-0.165	0.221	-0.638	0.307	90.27***	14	84.49%

(continues next page)

Tenure								
TMT	22	0.084	0.076	-0.074	0.242	126.62***	20	84.21%
Work	22	-0.08	0.08	-0.247	0.087	140.26***	20	85.74%
R&D	22	-0.016	0.098	-0.219	0.187	85.56***	20	76.63%
Mixed	---	---	---	---	---	---	---	---
Functional								
TMT	24	-0.072	0.075	-0.228	0.084	184.63***	22	88.08%
Work	24	-0.03	0.085	-0.206	0.145	298.84***	22	92.64%
R&D	24	0.299	0.088	0.117	0.481	59.58***	22	63.08%
Mixed	24	-0.055	0.16	-0.386	0.277	346.42***	22	93.65%
Education								
Level								
TMT	11	0.233	0.098	0.012	0.454	21.29**	9	57.72%
Work	11	-0.236	0.097	-0.455	-0.017	21.29**	9	57.72%
R&D	---	---	---	---	---	---	---	---
Mixed	11	0.01	0.238	-0.529	0.549	40.12***	9	77.57%

Note. k = total number of correlation coefficients meta-analyzed; r_z = corrected population correlation (sample-size weighted based ES on Fisher's z transformed correlation coefficients with significance test for ES=0; r_z = standard error of the corrected population correlation; *95% CI* = lower and upper bound of the 95% confidence interval; Q_E = homogeneity statistics; df = degree of freedom; I^2 = percentage of between-study variation due to heterogeneity.
***p<.01; **p<.05; *p<.1

With regard to the highly task-related diversity attributes of tenure, functional and educational background as well as education level heterogeneity, we predict in H3b the positive relationship between diversity and performance to be stronger in TMT and R&D teams in comparison to other team types.[10] Yet, the results are somehow mixed. While we observe the expected positive moderating effect of TMT teams, findings for R&D teams even report a slightly negative moderating effect. When it comes to functional background diversity, results are vice versa. Here, we find a strong and positive moderating effect in R&D teams, but a slightly negative effect in TMT. Fairly unexpected, reported findings on work teams and mixed teams depict a negative moderating effect on all three heterogeneity attributes. In contrast to our expectations, this indicates that diversity within work teams and mixed teams functions in line with similarity-attraction paradigm and social categorization theory. Overall, we are able to confirm hypothesis H3b only partially.

[10] There are no findings reported for educational background diversity due to insufficient observations.

2.6 Conclusion and Implications

Using meta-analysis, we study the influence of team diversity on team performance and productivity along a broad set of demographic attributes. These are understood to be either less task-related (age, gender, cultural background/ethnicity) or highly task-related (organizational tenure, functional and educational background, education level). In addition, we test for two potential moderating effects of the diversity-performance relationship – team size and team type.

When discussing the effects of homogeneity and heterogeneity in team composition on team outcome, diversity research mainly draws on two guiding research traditions: similarity-attraction paradigm/social categorization theory and information processing perspective/economic theory. The former tradition assumes heterogeneity to trigger social categorization, which may result in sub-group formation (in-group vs. out-group) and team conflict both assumed to be negatively associated with team performance. In order to avoid negative impacts on group performance, homogeneity in teams is stated to be preferable. Opposed to this, economic theory and information processing perspective argue that teams may benefit from heterogeneous cognitive resources, experiences or network ties – in particular when these are complements. As a consequence, heterogeneity in teams is reasoned to increase team performance. We use both theoretical concepts to derive the research hypotheses tested throughout this meta-analysis.

Core findings of the meta-analysis suggest negative population correlations of heterogeneity in less task-related attributes (gender and age) on team performance. This result confirms arguments stated in social categorization theory and similarity-attraction paradigm. Additionally, we detect a positive relationship between heterogeneity in highly task-related attributes (educational background) and team performance as expected by information processing perspective / economic theory. Moreover, we identify team size and team type to moderate the diversity-performance relationship to a large extent. In line with Stewart (2006), we are able to confirm the predicted moderating effects of team size: With increasing team size, the influence of diversity on performance decreases for less task-related attributes and increases for highly task-related attributes. Furthermore, we observe a positive moderating effect of TMT and R&D team type on almost all attributes irrespective of the task-relatedness. Fairly unexpected, we find the work team type to negatively moderate the diversity-performance relationship for all attributes, again irrespective of the task-relatedness. We attribute these findings either to compositional effects – TMT and R&D teams are assumed to be quite homogeneous along surface-level demography (e.g. Horwitz, Horwitz 2007, Devine 2002) – or to the very particular task requirements these teams face concerning creativity, decision making and problems solving.

Our findings offer profound implications for practitioners that might improve team cooperation and performance in organizational settings. Given the negative impact of age and gender heterogeneity on overall team performance, it seems to be recommendable to increase overall homogeneity on these attributes in teams within organizations. However, this mainly holds true for teams classified as working teams in this study (e.g. teams in the production line) since both TMT and R&D team types positively moderate age and gender heterogeneity. With regard to the positive impact of educational background diversity on overall team performance, one might consider interdisciplinary teams since these are found to outperform homogenous teams. Again, this mainly holds true for TMT and R&D teams because work team type as well as mixed team type negatively moderate this relationship. To put it another way, it seems reasonable to increase heterogeneity even along less task-related attributes in teams that require creative thinking, problem solving ability or decision making skills to fulfill their assigned tasks. This usually extends the cognitive resources of these teams and, therefore, improves team performance. In contrast to that, teams that follow standardized and routine tasks without the need to find innovative solutions are better off if team members are homogenous. These teams do not benefit from extended cognitive resources, but instead are likely to suffer from team conflicts that may arise through social categorization. Similar implications are offered regarding team size. Our results emphasize the advantageousness of large teams when highly task-related diversity is concerned. In other words, if a team requires creativity, problem solving ability or decision making skills to fulfill its assigned tasks, each additional team member may contribute complementary skills and knowledge. As a result, cognitive resources within the group increase and task performance may improve. Opposed to this, smaller teams are preferable if teams perform routine tasks since there is a negative moderating effect of team size for less task-related heterogeneity.

With this meta-analysis, we contribute to the diversity research in several regards. We provide an up-to-date meta-analysis on the diversity-performance relationship by focusing on studies published within the last twenty years, from 1994 to 2014. Meta-analysis seems particularly important in diversity research for two reasons. First, related experimental and empirical studies provide inconsistent and conflicting findings with regard to the productivity effects of a homogenous or heterogeneous composition. Hence, reviewing the literature qualitatively is highly unlikely to identify statistical patterns that may allow for universally valid conclusions. Instead, it seems necessary to quantitatively summarize the most important contributions to provide conclusions and implications that are generally valid – in particular in a field developing at such a vigorous speed. Second, primary studies are mainly built on small sample sizes that may raise doubts on the conclusions drawn from these contributions. In quantitatively analyzing these studies we might reduce existing doubts and contribute to the overall understanding of diversity. Furthermore, we consider only studies on work teams in real life organizations and leave out other team settings such as sport teams or student work groups. This seems necessary to derive business-oriented insights and evidence that are actually useful to enhance productivity within firms. In doing so, we contribute straightforward implications for practitioners on the advantageousness of team heterogeneity and homogeneity with regard to team performance. These insights may help to enhance team productivity and, thus, firm profitability.

Despite its profound contribution, this study entails some limitations. First, we are limited in our moderator testing on team size and team type. However, there is broad evidence on further potential moderators such as organizational culture and context (e.g. Riordan, 2000), group longevity (e.g. Pelled, Eisenhardt, Xin 1999) and task type in terms of complexity, autonomy and interdependency (e.g. Bowers, Pharmer, Salas 2000, Horwitz, Horwitz 2007, Webber, Donahue 2001). Unfortunately, we could not test for these moderators due to the fact that data on these issues is seldom reported in most primary studies. Second, the majority of primary studies summarized in this meta-analysis is based on cross-sectional data. This is somehow problematic since both theoretical assumptions (e.g. Harrison et al. 2002) and research findings (e.g. Pelled, Eisenhardt, Xin 1999, Watson, Kumar, Michaelsen 1993) suggest that timing matters in the relationship between diversity and performance. In other words, diversity seems to be a dynamic construct that may vary with group longevity and, therefore, should be studied with longitudinal rather than cross-sectional data. We call for future research to take these issues into consideration in order to arrive at a more detailed knowledge of the dynamics and moderating effects of diversity in work groups.

3 Employee Absenteeism: Determinants in the International Context

3.1 Introduction

In most industrialized countries employee sickness absence has become a significant trigger for high public and private expenses and probably will remain a central issue for future economic growth of both national economies as well as private organizations. In the US, the overall absence rate in 2013 totals to 2.9% (United States Bureau of Labor Statistics 2014), whereas in the European Union (EU-27 plus Norway) absence in 2010 varies between 0.8 in Italy to 7.7% in Norway with an overall mean of 3.8% (European Foundation for the Improvement of Living and Working Conditions 2010).[11] Detailed data for Germany reveal that employees missed a total of 567.7 million days at work due to sick leave in 2013. In other words, each employee has been absent from work for 15 days on average (BAuA 2015). With respect to these statistics, it comes as no surprise that employee absenteeism imposes very high economic costs for national economies and private organizations alike.[12] For the EU-27 plus Norway the total economic costs of employee absence are estimated to account for approximately 2.5% of the European Gross Domestic Product (GDP). For Germany, the costs imposed by employee absenteeism are estimated to amount to €59 billion in national production loss and even €103 billion in loss in gross value added (BAuA 2015). Given these numbers it becomes apparent that understanding the patterns of absenteeism may be crucial in order to enhance productivity and keep up with increasing global competition.

In response to the costly nature of absenteeism for both public institutions and private organizations, practitioners and academics are eagerly seeking for new insights on possible determinants of absenteeism in order to develop instruments that benefit both employees' health and employers' bottom-lines. This is even more relevant now than ever since recent and future demographic shifts in industrialized societies can be assumed to prolong the

[11] It is worth noting that international comparison of absence figures has proven to be difficult due to national differences in recording methodologies (e.g. maternity, child care) that limit overall comparability of absence data (European Foundation for the Improvement of Living and Working Conditions 2010).

[12] Costs imposed by absenteeism may be split up in direct and indirect costs (Martocchio 1992). Direct costs comprise all wages and statutory sickness benefits paid to employees on sick leave as well as expenses for replacement workers to fill in the vacancies. Indirect costs may arise in terms of productivity losses, costs for rescheduling and administration, reduced quality of service, safety risks and motivational issues for attendant employees (European Foundation for the Improvement of Living and Working Conditions 2010).

overall working life span (e.g. Grund, Westergaard-Nielsen 2008). On this account there is a great need for reducing existing workplace burdens that might provoke stress or illnesses in order to sustain employees' health in the long run. Detailed knowledge on the determinants of absence may help tackling these issues.

In general, absence can be defined as non-attendance at work when scheduled to (e.g. Kristensen et al. 2006). This idea allows to distinguish between involuntary (e.g. sickness) or voluntary (e.g. shirking) absence behavior (e.g. Johns 1997, Sagie 1998). In lack of a unique theory on absence, several attempts to conceptualize absenteeism have emerged over the years with the well-known process model of Steers and Rhodes (1978) being the most commonly cited. In their tradition, absence is usually understood as being subject to various personal, organizational, social and environmental variables that influence both motivation and ability to attend work. Much of the (personnel) economics work on absenteeism relates to either neoclassical labor supply models (Allen 1981a/b, Dunn, Youngblood 1986) or efficiency wage theory (Shapiro, Stiglitz 1984). Theory is backed up by experimental and empirical results on demographics (e.g. Voss, Floderus, Diderichsen 2001; Barmby, Ercolani, Treble 2002), group absence norms (e.g. Bradley, Green, Leeves 2007; Bamberger, Biron 2007), work satisfaction (e.g. Kristensen et al. 2006), working conditions (e.g. Dionne, Dostie 2007), national sickness benefit regulations (e.g. Frick, Malo 2008; Johansson, Palme 2005), employment protection (e.g. Riphahn, Thalmaier 2001; Ichino, Riphahn 2005) as well as economic environment and unemployment (e.g. Leigh 1985) – yet at some instances with inconsistent findings.

In this article, we empirically investigate why absence figures of blue-collar employees vary considerably between different production sites of a large corporation although employees face similar circumstances in terms of working conditions, manufacturing processes and final product at each facility. In particular, we are interested in the role of economic and social determinants in determining employee absence. Moreover, we examine the influence of worker characteristics and working conditions on absence. Using a hitherto unavailable data set covering 160 blue-collar work units at four international production sites of a large European automobile manufacturer, we offer a comprehensive analysis of potential determinants of absenteeism including:

- social peer influences (team size, team turnover, the share of temporary employed workers and the share of workers suffering health impairments)

- economic influences / incentives (sick pay regulations, employment protection, (un)employment and prosperity level)
- employee characteristics (age, gender, tenure and health)
- working conditions (shift system)

Following the data analyses, we conducted a total of eighteen interviews with on-site experts to further interpret our empirical findings. Experts include HR managers, line managers, worker representatives and shop floor staff.

The results presented in this paper indicate that absence at the respective corporation is a multifaceted phenomenon that only shows few statistical patterns that are valid at all four plants under observation. Instead, all determinants examined are differently affecting employee absence at each of the four production sites – details are discussed throughout chapter 3.5. In a nutshell, we observe a positive link between absence and social peer influences such as unit size and unit turnover. Furthermore, a high share of temporary workers increases absence of permanently employed workers while we do not find any absence effect of the share of workers suffering health impairments. The results regarding economic influence suggest strict employment protection laws as well as a favorable prosperity level to significantly increase absence, whereas findings for the national (un)employment situation remain inconclusive. Results for national sickness benefits legislation fail to reach statistical significance. With regard to worker characteristics, we observe unit absence to increase with the share of female employees. Additionally, we detect a positive link between age and both involuntary as well as voluntary absence. Evidence on the relation between absence and tenure as well as individual health is somewhat inconclusive. Eventually, results on working conditions imply that absence is higher in a two-shift and three-shift systems compared to the standard one-(day-)shift system. In chapter 3.6, practical implications for attendance management are considered based on the empirical findings and insights gained during expert interviews.

Our study contributes to the existing literature in several regards. First, we add to the growing research on determinants of absenteeism by offering new perspectives from inside an organization. In particular, we benefit from unique data that allows for comprehensive empirical testing. Second, we appear to be among the first researchers who are able to work with international data from within only one organization. Since all data is reported on company standards, we can neglect organizational and international differences in recording methodology that usually limit international comparison (e.g. European Foundation for the Improvement of Living and Working Conditions 2010). Third, our data allows us to examine three absence measures. Thus, we are able to empirically proxy for voluntary vs. involuntary absence. Furthermore, by using unit-level absence data we address a weakness of previous absence research that has focused almost exclusively on individual-level absenteeism and only account for unit-level effects in patches (e.g. Rentsch, Steel 2003). However, since absence is usually modeled as a social phenomenon (e.g. Steers, Rhodes 1978, Kaiser 1998) it should, by definition, be influenced by unit-level peers. Finally, our study adds to the growing insider econometrics approach that has been first introduced to personnel economics in seminal papers of Ichniowski, Shaw and Prennushi (1997) as well as Lazear (2000b).

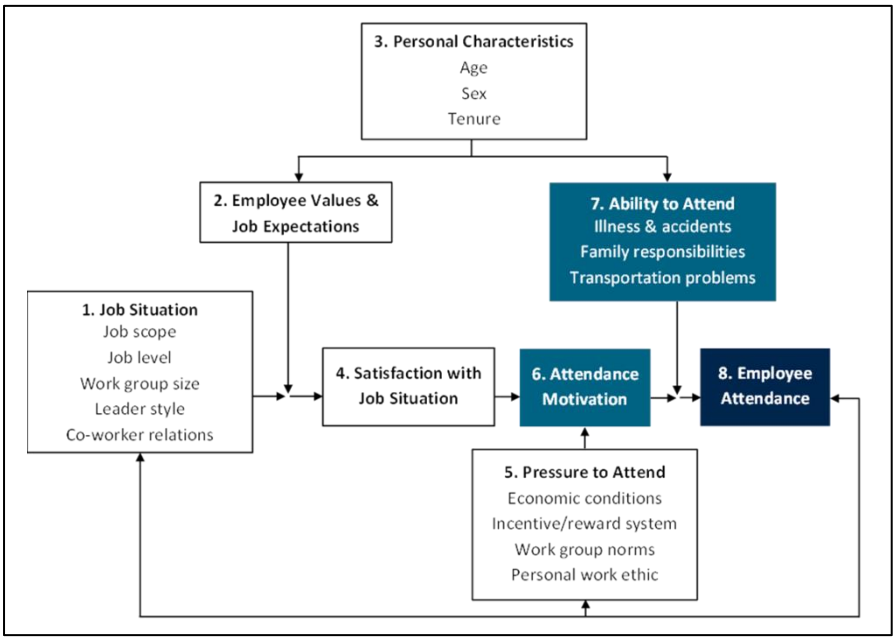
The remainder of this paper is organized as follows. The upcoming section provides both a discussion of the absenteeism concept as understood in this paper as well as a review of the existing absence research. Section three will present the data set and the peculiarities of the organizational setting under observation. Section four discusses the estimation strategy while section five presents the empirical results. Eventually, section six concludes and offers implications for both practitioners and future academic research.

3.2 Theoretical Framework

Research on absenteeism is multidisciplinary. It mainly originates from management literature but more recently arouse interest in (labor) economics and social psychology (Kaiser 1998). However, despite a substantial body of experimental and empirical evidence on absenteeism, a comprehensive understanding of the absence concept and its functioning is still missing. So far, research has merely agreed on a common definition of employee absence that includes all types of non-attendance at work when originally scheduled to, yet excluding scheduled absence agreed on with the employer such as holidays or flextime leaves (e.g. Kristensen et al. 2006). In other words, absenteeism is usually assigned to self-certified or medically-certified sickness absence and recognized as such by the employer (e.g. Whitaker 2001). In this study, we follow this definition of absence.

A "*conceptual breakthrough*" (Kaiser 1998, p. 81) in defining a theoretical framework of psychological absence research is the process model introduced in a now seminal article by Steers and Rhodes (1978). In their model, absence is understood as an individual decision based on the ability and the motivation to come to work (see Figure 3.1). In their model, the authors address weaknesses of previous contributions in questioning the assumptions of job satisfaction being the primary cause for absence (a comprehensive overview is given by Kaiser (1998)). In response, Steers and Rhodes (1978) model absence to be depending on both the motivation as well as the ability to attend work which in turn both are directly influenced by personal characteristics (e.g. age, gender, tenure). Additionally, the motivation to attend is further subject to the individual job situation (e.g. job scope, job level, work group size) as well as various attendance pressures (e.g. economic conditions, work group norms). The authors assume the ability to attend (e.g. sickness, family responsibilities) to moderate the effect of attendance motivation on absence. Later, the authors modified their first model into an advanced diagnostic model by including economic and social psychological factors that have not been considered in their initial model (Rhodes 1990).[13]

Figure 3.1: Steers and Rhodes (1978) Process Model of Employee Absence



Source: Illustration of Steers and Rhodes process model (Steers, Rhodes 1978, p.393).

In economic research, the understanding of absence differs from the psychological concepts. Much of the economic absence research relates to two main theories. First, neoclassical labor supply models assume absence to occur from individual day-to-day labor-

[13] Two alternative theories of absenteeism are discussed in the psychological absence literature. First, the withdrawal model which assumes absence to be withdrawal from unfortunate work circumstances (e.g. Hanisch, Hulin 1990, Porter, Steers 1973). In other words, absence may provide employees with stress-relief and allows them to return to work more productive (e.g. Bachler 1995). Second, the social-psychological approach that emphasizes the importance of absence norms within groups (e.g. Kaiser 1998, Rentsch, Steel 2003).

leisure choice decisions (e.g. Allen 1981a/b, Brown, Sessions 1996, Dunn, Youngblood 1986). The idea is simple: the individual utility is maximized when the marginal rate of substitution between income and leisure equals the wage offered by the employer. However, due to imperfect labor markets and job search costs, one can assume individuals to accept a job offer even if this constraint is not fulfilled at the contracted number of work hours. If at a given wage the contracted number exceeds the desired number of work hours, the individual has an incentive to be absent from work (Allen 1981a/b). Yet, this only holds true as long as the utility gains of being absent exceed the individual costs of absenteeism. These costs may include the gap between regular pay and sick pay. In other words, employees may use absence to maximize their individual utility (Dunn, Youngblood 1986). Therefore, absenteeism can be seen as "*a desirable nonpecuniary element of the compensation package"* (Allen 1981b, p. 207). Second, efficiency wage theory as discussed by Shapiro and Stiglitz (1984) assumes absence to arise from moral hazard and shirking. Therefore, absence may serve as a measure of individual effort level choice decisions and worker productivity and is widely used as such in the research (e.g. Ichniowski, Shaw 2013). The Shapiro-Stiglitz (1984) model is based on the assumption that the value of a present state (i.e. an employment relation) consists of the net current return as well as the expected future return. By being absent from work an employee risks dismissal and, as a consequence, the loss of the lifetime utility associated with the employment relation.

Both neoclassical labor supply and efficiency wage theory share the assumption that absence, at least partially, is an employee's individual decision under some constraints imposed by the employer. Hence, absence needs to be considered appearing either involuntary (e.g. as consequence of sickness or injury) or voluntary (e.g. to maximize individual utility). While involuntary absence spells are not expected to be influenced by motivation, voluntary absence can be modeled as utility maximization or shirking (e.g. Barmby, Sessions, Treble 1994, Brown, Sessions 1996, Chadwick-Jones, Brown, Nicholson 1973, Sagie 1998). Moreover, in the tradition of Steers and Rhodes (1978) absence can be interpreted as a social phenomenon with absence decisions never made solely by the individual, but instead being subject to certain constraints imposed by peers, subordinates, superiors and the organization's overall absence culture (e.g. Chadwick-Jones, Nicholson, Brown 1982; Rentsch, Steel 2003). Nevertheless, a great part of absence research focus exclusively on individual level absence (e.g. Kaiser 1998, Rentsch, Steel 2003).

Considering a potential sorting of determinants of absence variance we are loosely inspired by Kaiser (1998). He suggests absenteeism to be determined by personal characteristics and individual responses to influences stemming either from the work or non-work environment. Adding our own reading of the literature, we identify four fields of potential determinants of employee absenteeism: social peer influences in teams, economic influences / incentives, worker characteristics and working conditions. In what follows, we briefly discuss the determinants of absence along this classification in order to derive our main research hypotheses.[14]

3.2.1 Social Peer Influences in Work Teams

Turnover

It has often been pointed out in absenteeism research that absence variance is small within and large between groups (e.g. Xie, Johns 2000). Usually, this phenomenon is attributed to social absence norms that exist either among groups, organizations or cultural spheres (e.g. Bamberger, Biron 2007, Rentsch, Steel 2003). Absence norms as understood in this paper can be defined as "*set of absence-related beliefs, values and behavioral patterns that are shared among members of a work group or organizational unit*" (Gellatly, Luchak 1998, p. 1086). Although there is comprehensive empirical evidence on the link between peer absence and individual absence (e.g. Martocchio 1994, de Paola 2010, Ichino, Maggi 2000, Rosenblatt, Shapira-Lishchinsky, Shirom 2010), little is known on the social mechanisms behind this relationship (Johns 1997).[15] Since team members are rational utility-maximizers, we assume absence norms to be most likely in favor of the employees and, thus, increase absence. Now, consider a team whose members share a common belief about the legitimacy of absence. Within this team, the enforceability of the shared absence norm is subject to changes in the group composition since new workers are not aware of any existing absence norm or might not agree on participating (e.g. Miners et al. 1995). As a consequence, team-level absence can be assumed to decrease with team turnover. This argument leads us to hypothesis H1:

H1: "Absence on the unit-level decreases with turnover."

---

[14] We limit our review of the literature to those determinants that will actually be surveyed within this study. For a comprehensive overview on absence research see, for instance, review articles and meta-studies by Beemsterboer et al. (2009), Duijts et al. (2007) as well as Farrell and Stamm (1988).

[15] One possible argument is built on self-categorization (Turner 1987) and social identification theory (Taijfel 1982) since individuals prefer working with similar peers and, thus, adopt a particular peer behavior.

Unit size

Referring to neoclassic labor supply models as well as efficiency wage theory, one can assume that some workers stay away from work without being genuinely sick. Given this assumption, researchers often model absenteeism as a potential indicator of shirking (e.g. Barmby, Orme, Treble 1995, Riphahn 2004, Bradley, Green, Leeves 2007). Economic theory suggests that the share of shirking employees usually depends on the probability of being detected as well as the potential consequences once being caught (e.g. Alchian, Demsetz 1972). In this context, the effect of group size on shirking is straightforward: Shirking can be deterred by appropriate actions either by employers (e.g. organized monitoring, incentive design) or by co-workers (e.g. mutual monitoring, social sanctions).[16] However, any of these actions will be harder to enforce with increasing group size. For instance, the ability to mutually monitor co-worker behavior is getting more difficult with each additional team member and, as a result, the effectiveness of social sanctions that discipline workers who are caught shirking declines. Likewise, monitoring by supervisors is getting more complicated and/or costly with increasing team size (Baron, Kreps 1999). Furthermore, from Bandiera, Barankay and Rasul (2005, 2009) we know about the importance of social ties and familiarity among team members in determining productivity. The magnitude of worker familiarity, however, can be assumed to decrease with team size. Ultimately, one can predict shirking to be facilitated in larger teams due to reduced monitoring and less familiarity. These predictions are expressed in hypothesis H2:

H2: "Absence on the unit-level increases with team size."

Temporary workers

It is widely recognized in economic research that temporary workers are less absent than permanent workers. Yet, as soon as taken on a permanent contract they significantly increase their absence behavior (e.g. Bradley, Green, Leeves 2007, Engellandt, Riphahn 2005). This phenomenon is mainly attributable to incentivizing contract characteristics since temporary agents seek to qualify themselves for permanent contracts that are usually associated with better working conditions (e.g. Aronsson 1999, Paoli, Merllié 2001) and higher pay (e.g. Mertens, Gash, McGinnity 2007). However, it remains an open question how temporary agents influence their permanently employed co-workers with regard to absence behavior. Two potential arguments may be considered. First, assume that permanently employed workers are aware of the incentivizing nature of temporary contracts.

[16] On the importance of monitoring to reduce shirking in team settings recall Holmström (1982).

Being rational utility-maximizers, permanent agents might allow themselves a more “generous” absence behavior as they suppose their temporary co-workers to be incentivized to fill in the gap left by their own absence. This implies that a high share of temporary workers increases the absence of the permanently employed co-workers due to shirking. Second, as mentioned above, temporary employment is usually associated with worse working conditions. Suppose, for instance, that temporary agents are appointed to the most burdensome working conditions, e.g. by being excluded from job rotation. In turn, this would allow permanently employed workers to stay on less demanding work tasks and suffer less work-related strains. Following this argument, a high share of temporary workers might decrease permanent workers’ absence. Since we know that temporary agents are appointed on the same tasks and meet identical working conditions than permanent workers at the observed company, we believe the latter argument to be less convincing in our study design. Therefore, we base our thoughts on shirking behavior and hypothesize:

H3: “Absence of permanently employed team members increases with the share of temporary team members.”

Workers with health limitations

It is common practice at the studied organization that workers suffering from particular certified health limitations might be subject to temporary or permanent health impairments such as the prohibition to carry heavy loads or work overhead. As a result, these workers are exempt from performing tasks that would further deteriorate their health condition. Yet, it remains an open issue if the presence of employees with health impairments affects overall unit absenteeism. This question is mainly based on two considerations. First, it seems to be quite a straightforward argument that health-restricted employees are more prone to absence due to their physical (or mental) condition. Medical research largely emphasized the great significance of chronical and pre-existing health impairments in determining absence (e.g. Dewa, Lin 2000, Kessler et al. 2001). Second, the easier jobs in the team’s operations may often be occupied by workers with health impairments and, as a consequence, are excluded from unit-intern job rotation. In this situation, initially healthy employees might face additional burdens as they need to staff those workplaces that their team mates with health impairments cannot cover. These higher burdens may increase their own absence. Therefore, we expect a positive relationship between absence and the share of team members with health-impairments.

H4: "Absence on the group-level increases with the share of team members with health impairments."

3.2.2 Economic Influences / Incentives

Sickness benefits

It is a stylized fact in economics that the institutional framework determines employees' absence behavior. Early evidence is given by Buzzard and Shaw (1952) who state that absence increase with higher sick pay replacement rates as the costs of absence decrease for employees. A corresponding labor supply model is offered by Brown and Sessions (1996). They argue that higher sickness benefits raise incentives of workers to be absent from work. These assumptions are supported by Frick and Malo (2008). Based on the European Survey of Working Conditions they observed a significant increase in absenteeism with more generous sickness benefits within the EU-14. Over the years, several EU governments initiated changes to national sickness legislation in order to address the burdens of sick pay for social security systems and employers. Absence research often benefits from these legislative changes that allow for natural experiments, e.g. Johansson and Palme (2002, 2005) as well as Voss, Floderus and Diderichsen (2001) for Sweden and Puhani and Sonderhoff (2010) as well as Ziebarth and Karlsson (2010, 2013) for Germany. These studies provide broad support of the positive link between statutory sickness benefit regulations and employee absence. We, therefore, predict:

H5: "Absence will increase with more generous sickness benefit regulations."

Employment protection

The same line of argument holds true for employment protection. Focusing on a legislative change in Sweden, Olsson (2009) observes that a less favorable labor protection law in terms of seniority dismissal protection significantly decreases absence behavior in organizations that are affected by the change. Based on German data, Riphahn and Thalmaier (2001) find that absenteeism increases after probation periods end and mandatory labor protection sets in. Further evidence is given by Ichino and Riphahn (2005) who observe the same behavior for Italian bank employees. Bradley, Green and Leeves (2012) report a significant raise in absenteeism when temporary workers are taken on permanently. They argue that the low employment protection associated with temporary employment incentivizes workers to show high effort in terms of low absence. As soon as labor protection applies to their employment relation, the formerly temporary agents reduce effort and no longer refrain from shirking. This argument leads to hypothesis H6:

H6: "Absence will increase with strict employment protection legislation."

Unemployment and prosperity level

In addition to the legislative framework, the economic environment and unemployment rates are often linked to absence behavior beginning with a seminal paper by Leigh (1985). Using data from the time of the US recession in the 1970s, he observed employees to refrain from absenteeism when unemployment is high as they assume employers to first lay off workers who have proven absence-prone. Askildsen, Bratberg and Nilsen (2005) report similar findings for the Norwegian labor market: An increase in local unemployment rates significantly decreases absence. This result even holds when controlling for workforce composition which is often claimed to be responsible for cyclical variation in absenteeism. Knutsson and Goine (1998) observe the link between absence and unemployment for Swedish males. With regard to the economic situation, Virtanen et al. (2005) observed for Finish public sector employees that a constantly poor local economy has a decreasing impact on self-certified sickness. Audas, Goddard (2001) confirm the link between business cycle and absenteeism for the US. Thus, we predict:

H7: "Absence will decrease with increasing unemployment."

H8: "Absence will increase with prosperity level."

3.2.3 Worker Characteristics

Non-work related worker characteristics (gender, age)

The link between non-work related worker characteristics (i.e. gender and age) and employee absenteeism is well established. While for gender it is a stylized fact that females exhibit higher sickness absence than males (e.g. Barmby, Ercolani, Treble 2002, Mastekaasa 2000), results for age are less conclusive (e.g. Rhodes 1983).

According to Bekker, Rutte and van Rijswijk (2009), there are three main reasons for gender differences in absenteeism. First, due to physical differences women suffer more from reproduction-related health issues such as pregnancy and menstruation (e.g. Alexanderson et al. 1996, Sydsjö, Sydsjö, Alexanderson 2001). Second, females face other daily obligations than males such as the double burden of job and family since child care is traditionally more associated with women (e.g. Åkerlind et al. 1996, Bratberg, Dahl, Risa 2002, Vistnes 1997). Yet, recent research often fails to link child-care to absence (e.g. Mastekaasa 2000, Erikson, Nichols, Ritter 2000). Third, male and female occupations usually are clearly separated based on the socio-cultural concept of gender role orientation (e.g. Messing et al. 1998). This involves varying job strains (e.g. Laaksonen et al. 2010), and different absence norms in female-dominated occupations (e.g. Mastekaasa 2005). Thus, we hypothesize:

H9: "Absence on the group-level increases with the share of female group members."

The Steers and Rhodes (1978) model associates age with both involuntary and voluntary absence. On the one hand, the relationship between age and involuntary absence seems to be direct as age-related factors such as illnesses and accident risk negatively influence the ability to attend (e.g. Rhodes 1983). It is widely accepted that physical fitness and overall health condition are negatively related to age. As a result, older workers suffer from diminishing physical resources to cope with work-related and non-work related stressors and are, therefore, more prone to work-related illnesses (e.g. Ilmarinen 2001, Ng, Feldman 2013). Meta-analytic evidence confirms the positive relationship between age and sickness absence (e.g. Ng, Feldman 2008). This led us to assume:

H10: "Involuntary absence as proxied by absence spell duration increases with unit mean age."

On the other hand, voluntary absence is indirectly influenced by values, expectations and satisfaction that are moderated by age (e.g. Rhodes 1983). Findings on voluntary absence measures (e.g. absence frequency) indicate a negative relationship between age and absence (e.g. Chadwick-Jones, Nicholson, Brown 1982, Leigh 1986). In other words, older employers are less likely to shirk without being genuinely sick. This behavior is often linked to higher social and financial duties that usually increase with age. Meta-analytic evidence emphasizes the negative influence of age on voluntary absence (e.g. Hackett 1990, Martocchio 1989). Based on these considerations we hypothesize:

H11: "Voluntary absence as proxied by absence frequency decreases with unit mean age."

Work-related demographics (tenure as proxy for work experience)

Evidence on the tenure-absence relationship is somehow inconclusive as a positive (e.g. Barmby, Ercolani, Treble 2002, Tompa, Scott-Marshall, Fang 2008, Ng, Feldman 2013) as well as a negative relationship is empirically supported (e.g. George 1989, Nicholson, Brown, Chadwick-Jones 1977). In a meta-analysis, Hackett (1990) even failed to confirm any tenure effect at all. However, there are reasonable arguments for the existence of both a positive and a negative tenure effect. On the one hand, absence can be assumed to decrease with tenure since more experienced workers have gained more (firm-specific) human capital. This may reduce the individual risk of injuries or illnesses since workers are familiar with the work-related risks and hazards (e.g. Breslin, Smith 2005). Moreover, we expect senior employees to have developed more elaborated strategies to cope with work-related stressors. On the other hand, absence can be assumed to increase with seniority since tenure is highly correlated with age and, thus, any tenure effect might just represent the age effect as discussed above (e.g. Barmby, Ercolani, Treble 2002, Gordon, Johnson 1982). Furthermore, seniority often increases job security that may allow workers to be absent without having to fear dismissal (e.g. Barmby, Ercolani, Treble 2002, Tompa, Scott-Marshall, Fang 2008). In the organizational setting at hand, we can neglect reasons of job security since the company imposes itself with strict employment protection regulations without differentiating between newly hired or veteran workers. Thus, we believe tenure to be valid proxy for work experience and hypothesize:

H12: "Absence decreases as work experience increases."

Acute health condition

Acute health impairments such as influenza and influenzalike illnesses (ILI) are among the top reasons for sickness absence and may account for 10% to 12% of total absenteeism (e.g. Keech, Scott, Ryan 1998).[17] Days of work lost due to annual ILI pandemics are found to vary between 0.3 to 5.9 days depending on the database (e.g. Keech, Scott, Ryan 1998, Tsai, Zhou, Kim 2014). In total, economic and social costs of ILI were estimated to amount up to $87.1 billion in the US in 2003 (e.g. Molinari et al. 2007). Although often advised only to children and older persons, ILI vaccination could help employers to significantly reduce cost imposed by ILI-induced absenteeism (e.g. Akazawa, Sindelar, Paltiel 2003). In light of these facts, we hypothesize as follows:

---

[17] In addition, there is comprehensive empirical work, which is not emphasized over the course of the present study, on the absence effect of chronical health condition (e.g. Dewa, Lin 2000, Kessler et al. 2001), obesity (e.g. Colditz 1999, Finkelstein, Fiebelkorn, Wang 2005) and addictive disorders such as smoking (e.g. Leigh 1995, Lundborg 2007) and alcohol abuse (e.g. Norström 2009, Bacharch, Bamberger, Biron 2010). With regard to acute health issues we focus exclusively on ILI. Certainly, we are aware of other short-term health impairments such as lower back pain (e.g. Dagenais, Caro, Haldeman 2008, Maetzel, Li 2002). Moreover, we exclude mental disorders and psychological problems since company representatives name ILI (and long-term musculoskeletal disorders) to account for the majority of work days lost.

H13: "Absence increase with acute health impairments."

3.2.4 Working Condition

Shift work

Non-standard working hours in terms of shift-work and night shifts have become a necessity to keep up productivity and business competitiveness (e.g. Costa 2003). In Germany, for instance, rotating shift work is affecting 15% of the total workforce and is applied at 42% of all manufacturing companies (Jirjahn 2008). It is widely recognized that shift work is a risk factor for negative physical and mental health outcomes as well as sleeping problems (e.g. Koller 1983, Akerstedt 2003, Knutsson 2003, Knutsson, Bøggild 2010). Furthermore, working shifts is detrimental to employees' social well-being since shifts are usually concurrently scheduled to social and family activities (e.g. Jansen et al. 2004). In consequence, there is a substantial body of empirical evidence associating both health and social burdens of shift work with increasing absence (e.g. Slany et al. 2014, Morikawa et al. 2001, Fekedulegn et al. 2013).[18] Since shift work is applied on a large scale at the company under observation, we consider the following hypothesis:

H14: "Absence is higher in the presence of shift work and night shifts."

All in all, we identified fourteen hypotheses that will be tested throughout this study. In the following paragraph we will present the data set and discuss the particularities of the organizational setting under observation.

3.3 Data Set and Descriptive Statistics

In order to study the determinants of employee absence in shop floor settings, we use a unique panel data set with team-level information on sickness absence of blue-collar employees at four European production sites of a large automobile manufacturer. Two of the four locations are low volume production sites for luxury cars. These plants are relatively small and employees operate at a moderate cycle time. One of these plants is located in Germany (GER), the other one in the United Kingdom (UK). The other two locations are large volume car production sites with employees operating at a short cycle time. These production sites are located in Germany and Spain (ESP), respectively (see Table 3.1 for an overview of plant characteristics).

[18] We are aware that there are studies observing inconclusive or no significant relationships between shift work and absence (see Merkus et al. 2012 for a review).

Team-level absence data includes information on the overall absence rate, the mean duration of absence spells as well as absence frequency. In order to investigate the determinants of absence, we match the available absence data with broad information on social peer influences, economic influences, worker characteristics and working conditions. All proprietary data is reported from company records on the work unit level (i.e. teams). Data comparability is guaranteed since all data share a common reference (the team-level) and identical reporting standards, e.g. recording of absenteeism data is based on similar group standards at all four production sites. In order to test for hypotheses H5 to H8 we add data gathered from external sources such as information on sick pay legislation, unemployment rates and economic environment. All variables used throughout this article are discussed in the next section in more detail. Our data covers an extended observation period of thirty-six consecutive months from January 2011 to December 2013. We chose monthly reporting intervals over weekly or daily intervals since this procedure levels out short-term effects and more or less identifies variables that are relatively stable over time (e.g. Harrison, Martocchino 1998). The data set covers information on a total of n=160 organizational work units (with n=2,152 employees as of December 2013) and n=5,673 unit-month observations. In general, a work unit can be defined as a group of blue-collar employees that work together as shop-floor teams either in the press shop, the body shop, the paint shop or the assembly line.[19] Within each team, job rotation is applied to level out the strains and burdens associated with some work places.

Teams were selected randomly by company representatives. Only teams that meet the following criteria were considered. First, teams had to exist for the complete observation period of thirty-six months. At this point stability is needed since at some sites teams were dissolved, consolidated or newly found due to organizational reasons (e.g. new products, new machinery etc.).[20] Yet, variance in personnel structure within teams is highly appreciated. Second, due to data privacy regulations, teams required reported data of five or more members for each month. Regarding the selection process, one might be concerned about a selection bias to be present. However, in our context, we can neglect this issue since all teams included in the data quite accurately represent the overall blue-collar work forces at the respective production sites in terms of absence, member characteristics and work tasks.

[19] Due to organizational reasons not all plants employ all four stages of automobile manufacturing.

[20] We have to allow two exceptions to this rule in order to reach sufficient sample size for both luxury car producing plants. At the one plant, two units were newly found during the course of the year 2011. At the other plant, four units were consolidated into two during the year 2013. However, this should not be a great concern in our data set since the teams did not differ systematically from other teams.

Table 3.1: Summary Statistics – Data Set Composition

Plant	*Country*	*Product*	*Cycle time*	*Organizational units covered by data set*	*Employees covered by data set (Dec. 2013)*
A	GER	Luxury	Moderate	6	136
B	UK	Luxury	Moderate	8	281
C	ESP	Volume	Short	104	1,245
D	GER	Volume	Short	42	490
Number of units: 160 Total number of unit-month observations: 5,673*					

* Missing values at some instances due to organizational changes

As already mentioned absence data is matched with information on social and economic influences, worker characteristics and working conditions. Some of the summary statistics are presented in Table 3.2 for each plant. We find age to be distributed relatively homogenous between the teams at each plant since standard deviations do not exceed 2.5 years. Concerning the gender distribution we find both luxury car production plants to have a very homogeneous labor force consisting mainly of males (97.02% and 97.74% males, respectively), while both volume car production sites employ a significant share of female workers (78.4% and 89.98% males, respectively). This might be reason to labor supply necessities. Unit size is quite similar at both luxury car producers with means of 32.8 and 35.0 employees, but differs largely from the average unit size at both volume car producers (12.1 and 11.9 employees). This seems reasonable since volume car production is usually more subdivided into small production segments than luxury car production. The descriptive absence data already reveal that some plants (B and D) show absence rates below or only slightly above the international average of 3.8% within EU-states (European Foundation for the Improvement of Living and Working Conditions 2010). In contrast to that, the other two plants (A and C) far exceed international reference values. This notion is supported by significant variances between plants concerning absence rate ($F(2, 4136)=43.95$, $p=.000$), mean absence duration ($F(2, 4137)=3.86$, $p=.021$) and absence frequency ($F(2, 4133)=429.56$, $p=.000$). More than one-third of all unit-month observations display an absence rate of zero (34.82%).

Table 3.2: Summary Statistics – Means and standard deviations

Plant	***Product***	***Unit size (in members)***	***Age (in years)***	***Gender (in % of males)***	***Tenure (in years)***	***Temporary workers (in %)***	***Divisions***	***Absence rate (in %)***	***Absence spell duration (in days)***	***Absence frequency****
A	Luxury	32.81 (8.31)	35.12 (1.82)	97.02 (2.87)	5.62 (.97)	14.84 (11.14)	Assembly	7.03 (3.44)	5.64 (2.38)	.2827 (.1115)
B	Luxury	35.00 (6.79)	43.33 (2.40)	97.74 (3.02)	16.50 (2.55)	13.25 (11.83)	Body, Paint, Assembly	2.73 (2.53)	5.01 (3.55)	.0844 (.0586)
C	Volume	12.10 (4.22)	37.43 (2.33)	78.40 (17.93)	10.99 (3.31)	n/a	Press, Body, Paint, Assembly	4.17 (5.19)	6.10 (6.91)	.0873 (.0925)
D	Volume	11.91 (3.30)	41.46 (3.45)	89.98 (12.35)	13.02 (2.36)	9.84 (17.04)	Press, Paint, Assembly	6.25 (5.45)	7.07 (6.81)	.1844 (.1512)

Number of units: 160
Total number of unit-month observations: 5,673**

* Absence frequency is calculated following Hensing et al. (1998) by dividing the number of absence incidents by unit size.
** Missing values at some instances due to organizational changes.

Our study design has numerous advantages in addressing issues raised by previous research on absenteeism. First, starting with early work by Chadwick-Jones, Nicholson and Brown (1982) absence is not solely modeled as an individual phenomenon but as being subject to social peer influences. Given that within-unit absence varies only little whereas between-unit absence varies widely one can assume unit-level influences to play an important role in determining absence (e.g. Rentsch, Steel 2003, Harrison, Martocchino 1998). Thus, it seems surprising that relatively little work is done on absenteeism at the group level (e.g. Rentsch, Steel 2003) with only few exceptions (e.g. Xie, Johns 2000, Kristensen et al. 2006). We address this gap in absence research by offering new insights on sickness absence at the team level. Second, absence research mainly relies on self-reported sickness data, e.g. gathered via questionnaires, since register data is often unavailable to researchers. However, the response sensitivity of self-reported data in comparison to register data varies from high (e.g. 91% (Voss et al. 2008), 88% (Burdorf, Post, Bruggeling 1996)), to medium (e.g. 82% (Aguis et al. 1994)) and low (e.g. 55% (van Poppel et al. 2002)). In line with the latter, Johns (1994) finds employees to be absent actually twice as much as self-reported.[21] Since we are able to use register data from company records we can neglect concerns about the reliability and validity of our data. Third, since our data set includes four production plants from Germany, Spain and the UK we are able to control for varying institutional frameworks across these countries. As discussed earlier, regulations for sickness benefits and employment protection are found to significantly determine absence behavior in the international context (e.g. Frick, Malo 2008, Riphahn, Thalmaier 2001). However, international comparability of absence statistics usually proves difficult due to widely varying reporting standards for specific cases such as family care or maternity (European Foundation for the Improvement of Living and Working Conditions 2010). For this study, we can dispel any doubts about international comparability since all absence data is recorded based on company standards. Finally, we are able to apply multiple measures of absence as dependent variables to account for different aspects of absenteeism such as voluntary and involuntary absence. All absence measures used in this study are discussed in detail in the following section.

[21] Moreover, accuracy of self-reported sickness absence is found to decrease with time (Severens et al. 2000) and with increasing number of sickness days (Ferrie et al. 2005, Johns 1994).

3.4 Estimation Strategy

Our empirical analyses are based on three different absence measures to consider absenteeism from various perspectives. We use absence KPIs reported on company recording standards. The most important absence figure is a unit's monthly absence rate. It is defined as the percentage of time originally scheduled for work within the unit missed due to self-certified and medically-certified sickness. All company-intern discussion on absence is entirely centered on this measure. Absence rates vary across units and locations, the overall statistics are presented in Table 3.2. Moreover, we follow Hensing et al. (1998)[22] in constructing further absence measures since working with register data does not allow for a clear separation between involuntary and voluntary absence. We, therefore, apply two monthly proxies for both absence types proposed among others by Sagie (1998). First, one can assume involuntary absence spells to be usually longer since sicknesses and injuries are seldom cured completely after one or two days. Hence, we calculate the mean duration of sick leaves of a team rounded to integer values (Hensing et al. 1998). Second, voluntary absence is widely recognized as short-lasting but more frequent. That is why frequency measures serve well as proxy for voluntary absence (e.g. Chadwick-Jones et al. 1971). We follow Hensing et al. (1998) in calculating a frequency measure by dividing the number of absence spells of a given team by the number of persons in that particular unit. Still, a definitive separation between involuntary and voluntary absence remains unclear since both proxies might include the other absence behavior as well.

In general, our estimations are based on the following models with *ABS* representing the three absence measures discussed above as dependent variables:

$$ABS = \alpha + \beta PEER + \gamma ECON + \delta WORK + \varphi COND + \vartheta CONT + \varepsilon$$

where α is the constant, *PEER* is a vector of social peer influences on the team level, *ECON* is a vector of economic influences / incentives, *WORK* is a vector of worker characteristics, *COND* is a vector describing working conditions and *CONT* is a vector of further controls, β, γ, δ, φ and ϑ are the estimated coefficients and ε is the error term. All variables are discussed in more detail below.

[22] Hensing et al. (1998) propose frequency, length, incidence rate, cumulative incidences and duration of sick-leave as measures of absenteeism. Although we are not able to construct all measures for our sample due to unavailable or missing data, we use their frequency and duration measures.

Social peer influences (PEER)

We model social influences of workers' peers to study their impact on employee absence. One important peer influence is focusing on group absence norms (e.g. Bamberger, Biron 2007, Rentsch, Steel 2003). These norms might change with turnover (Miners et al. 1995). With regard to hypothesis H1 we proxy for turnover by adding all new arrivals to and all exits from the team in order to create one overall turnover variable. The idea is simple: we use this quite intuitive measure of variance in team composition since absence norms should be influenced by both new and leaving workers alike. In addition, we address research on shirking in team settings (e.g. Dunn, Youngblood 1986) by stating that with increasing team size absence should be more pronounced due to facilitated shirking options (H2). To incorporate this idea in our estimations we include monthly team size for all units. We measure team size as the simple headcount of team members in a given month. As expressed by hypothesis H3, the share of temporary workers might increase the absence behavior of permanent workers. In order to control for this argument, we include the percentage of temporary workers in a unit in our estimations. Unfortunately, plant C did not report any information on the share of temporary workers. In hypothesis H4 we state that the share of team members with health impairments should increase unit-level absence for two reasons. First, most obviously these workers can be expected to be more prone to absence. Second, healthy workers might suffer additional workload. We take this hypothesized relationship into account by including the percentage of team members that have any kind of certified health impairment in a given month.

Economic influences / incentives (ECON)

In order to study potential economic influences and incentive effects on employee absence, we extend the company data set with environmental variables identified by previous research to determine employee absenteeism. At this point, we benefit from the international context of our study since we are able to compare different social security systems and employment protection laws as well as varying national economic environments. This allows us to consider potential economic influences along the four main topics introduced in H5 to H8. First, as proposed among others by Riphahn and Thalmaier (2001) as well as Ichino and Riphahn (2005) employment protection may have a significant impact on employee absence behavior. To account for varying legislative settings in Germany, Spain and the UK we incorporate the OECD's Indicators of Employment Protection (OECD 2014a). Three synthetic indicators rate the strictness of a country's regulations on individual and collective dismissal as well as on fixed-term and temporary work agency contracts on a scale from weak=0 to very strict=6. Second, we consider suggestions stating that absenteeism increases with more generous sickness benefits (e.g. Frick, Malo 2008, Johansson, Palme 2002, 2005) in following an approach proposed by Frick and Malo (2008). They use the so-called MISSOC reports (Mutual Information System on Social Protection) published by the European Commission that describe national sickness benefits concerning their coverage, waiting period, maximum duration and replacement rate (European Commission 2014a). We convert the report's textual descriptions into a scale from weak=0 to generous=6 sickness benefits in order to match the OECD's Indicator of Employment Protection scales. Third, following previous evidence (e.g. Leigh 1985, Askildsen, Bratberg, Nilsen 2005) we expect high unemployment rates to significantly influence employee absence behavior. Hence, we control for national (un)employment in two ways. On the one hand, we use seasonally adjusted monthly national unemployment rates. On the other hand, we include the share of persons employed in manufacturing (NACE Classification D) in percent of total employees to model workers outside options. Data is based on quarterly Eurostat statistics for both unemployment (Eurostat 2014a) as well as employment information (Eurostat 2014b). Finally, we take into consideration research stating a link between national economic situation and employee absence behavior (e.g. Virtanen et al. 2005, Audas, Goddard 2001) by incorporating Worldbank information on annual percentage changes of GDP by country (Worldbank 2014).

Worker characteristics (WORK)

To account for the influence of worker characteristics on absence we chose four main potential determinants of absenteeism. On the one hand, we include the classical non-work-related demographic features of age and gender to test hypotheses H9 to H11. First, age is included as unit mean age. Although the measure of age on the group level seems somehow not ideal since similar age values in groups may be caused by totally different age distributions (e.g. Kristensen et al. 2006), we believe that controlling for the coefficient of variation (CV) of age should level out potential biases. Second, we use the share of males to model the gender distribution within a team. On the other hand, we aim at addressing potential influences of work-related individual demographics on absence. We, therefore, use unit mean tenure and the respective CV as proxies for professional experience. Following this logic, a high mean tenure indicates units with workers that have accumulated high firm-specific human capital and, thus, should have better strategies to cope with work-related stressors. Although age and tenure are highly correlated (0.698), results do not change when leaving tenure out from the estimations. Therefore, it seems that age and tenure variables are actually measuring two different phenomena. This is in line with among others Barmby, Ercolani and Treble (2002) who observe a tenure effect on absence even when age is controlled for. Unfortunately, CVs were calculated neither for age nor for tenure from plant D. Eventually, we address research on acute health condition with influenza and ILI being among the most stated reasons for staying away from work (e.g. Tsai, Zhou, Kim 2014, Keech, Scott, Ryan 1998).[23] As suggested in hypothesis H13, we control for ILI and influenza by using influenza recordings from the World Health Organization (WHO). These statistics assess influenza activity from no activity to widespread outbreak on a national level (WHO 2014). We convert weekly textual evaluations to monthly ordinal scales of no activity=0, sporadic=1, local outbreak=2, regional outbreak=3 and widespread outbreak=4.

Working conditions (COND)

In H14 we address research stating that shift work usually increases absence due to higher burdens concerning sleeping habits, disordered diurnal rhythm and complicated family and social life (Slany et al. 2014, Morikawa et al. 2001, Fekedulegn et al. 2013). By taking a normal seven hour one-shift system (only morning shift) as reference category, we model additional shifts based on a two-shift (morning and day shift) or three-shift systems (morning, day and night shift).

Control variables (CONT)

In order to further exclude confounding effects on employee absence behavior, we add further controls to the company data set. First, we address evidence on the influence of weather on absenteeism (e.g. Shi, Skuterud 2015) by including information on monthly mean temperature, precipitation and sun hours.[24] Second, we use month and year dummies to control for potential seasonality or other timing effects of absence.

---

23 Yet, we do not have information on the individual health condition of employees or work units.

24 Since we aim to ensure data comparability by relying on single-sourced data provided from the international network of the German Meteorological Service (DWD 2014), we only have a limited choice of meteorological stations close to the plants' locations. Therefore, weather information was gathered at meteorological stations 130km away from sites at maximum. We do not assume this to be a problem.

3.5 Empirical Results

As a first step, we conduct one-way ANOVAs in order to check for significant mean variance in absence rates between and/or within organizational units. While we find between-unit variance in absence rates to be highly significant at all plants (plant A ($F(5, 192)=6.56$, $p=.000$), plant B ($F(7, 256)=8.60$, $p=.000$), plant C ($F(103, 3564)=9.67$, $p=.000$) and plant D ($F(41, 1419)=4.28$, $p=.000$)), we observe within-unit absence to vary significantly at only 69 out of 160 units over time (43%). This phenomenon is in line with previous research that found absence variance to be great between units but smaller within units (e.g. Harrison, Martocchio 1998, Xie, Johns 2000).

In response to this finding, we chose fixed effects regression models to be preferable over random effects models since there seems to be some group specific effects that remain constant over time. Fixed effects may be of psychological nature but may also be attributable to varying workplace characteristics that might increase ergonomic burdens of work. We, therefore, estimate fixed-effect models on all three absence measures to test for potential determinants of employee absenteeism.[25] In this process, we follow a two-staged estimation strategy. On the one hand, we pool data of two plants (both volume car production sites and both luxury car production sites, respectively) and apply pair-level estimations by using interaction terms for economic and social peer influences. This process seeks at avoiding potential biases arising from the fact that variables may not only vary between plants but may determine absence behavior at each plant in different ways. On the other hand, we apply plant-level estimations for each location independently to further control for plant-specific effects that might get lost in pair-wise regressions. For all estimations we use various model specifications including different sets of explanatory variables to control for the robustness of our findings.

Subsequent to the empirical testing, we conducted a total of 18 on-site expert interviews at both international sites (B, C) in order to double-check our own experience from national plants. Experts are working in different organizational fields including HR experts, worker representatives, shop floor managers and shop floor staff. Their operational experience and practical knowledge helped us to interpret our findings comprehensively and derive implications that are feasible and realizable in organizational implementation.

---

[25] Other researchers often use count data methods such as (zero-inflated) poisson or negative binomial regressions to study absence duration. Unfortunately, our duration data represents a unit's mean and, therefore, does not meet the integer value assumption necessary for both regression models.

In the following, we start the detailed discussion of our findings by first presenting estimation results on the effects of social peer influences on worker absence (H1 to H4). We then go on to present our findings with respect to economic influences / incentives (H5 to H8), worker characteristics (H9 to H13) and working conditions (H14). Within each paragraph we discuss pooled pair-level as well as plant-level results and further distinguish our findings along the three main dependent variables absence rate, absence spell duration and absence frequency. Overall findings suggest that determinants of absence vary by plant. In other words, we only find few patterns in the data that are common at all four plants. Main results of pooled interaction estimations are reported in Tables 3.3 to 3.5, main results of plant-level estimations are reported in Tables A.1 to A.3 in the appendix. Estimations in all tables are numbered for a better orientation and referred to throughout the text. Due to reasons of brevity we focus on presenting the main results in all tables.

Social peer influences (PEER)

Starting with peer effects, we have analyzed four potential determinants: turnover as a proxy for group norms, unit size as a proxy for shirking, the share of temporary workers and the share of workers with temporary or permanent health impairments. Results will be discussed in the specified order. In hypothesis H1, we predict absence to decrease with turnover because the enforceability of group norms – that we believe to increase voluntary absence – may be limited with variance in team composition. Yet, we find only weak support for this argument as the relationship between turnover and absence spell duration is only significant at plant D (estimation VI-4). However, since duration measures are usually used to proxy for involuntary absence, we refrain from interpreting this finding in support of H1. Instead, plant level estimations for location D (V-4) as well as interaction term estimations for locations C and D (III-4, -5, -6) indicate that high turnover increases absence frequency. Results for absence rate do not reach statistical significance. The great majority of local experts confirmed the empirical findings as they observe absence to be usually higher in units with high employee turnover. According to their experience, this behavior is attributable to employees feeling unsettled and uncertain rather than to changing group norms. Moreover, turnover and absenteeism are both identified to be strategies workers may choose in response to individual job dissatisfaction (theory of exit, voice and loyalty, e.g. Hirschman 1970, Farrell 1983). Thus, high turnover may reduce the workers' need of other responses to job dissatisfaction such as absence. As a consequence, absence can be assumed to be low in high turnover situations. In total, we have to reject hypothesis H1.

Table 3.3: Fixed-Effects Interaction Terms on Absence Rates (I)

	luxury sites						volume sites					
	I-(1)		I-(2)		I-(3)		I-(4)		I-(5)		I-(6)	
VARIABLES	**Plant A (GER)**	**Plant B (UK)**	**Plant A (GER)**	**Plant B (UK)**	**Plant A (GER)**	**Plant B (UK)**	**Plant C (ESP)**	**Plant D (GER)**	**Plant C (ESP)**	**Plant D (GER)**	**Plant C (ESP)**	**Plant D (GER)**
Mean age (in years)	-0.949	-0.611	-0.206	0.703*	-0.396	-1.512	-0.476	1.487	0.021	0.501**	1.601	0.257
	(7.203)	(2.612)	(0.494)	(0.385)	(7.600)	(2.699)	(1.501)	(1.908)	(0.163)	(0.223)	(2.635)	(2.762)
Mean age² (in years)	0.00949	0.0170	---	---	0.00260	0.0193	0.0059	-0.0134	---	---	-0.022	0.00293
	(0.0994)	(0.0287)			(0.105)	(0.0299)	(0.0201)	(0.0236)			(0.035)	(0.0338)
Share of males	-19.74	5.273	-17.50	7.464	-15.52	-0.192	-2.78*	-0.297	-6.988***	0.972	-6.751***	-0.364
(in %/100)	(18.96)	(8.150)	(15.06)	(7.947)	(20.10)	(9.218)	(1.429)	(4.916)	(2.406)	(6.67)	(2.342)	(6.735)
Mean tenure (in years)	2.533	-2.223*	-3.027**	-0.344	-0.967	-1.472	0.0492	0.5573	0.2374	-0.5294	0.1902	1.457
	(3.193)	(1.223)	(1.378)	(0.226)	(4.925)	(1.015)	(0.4074)	(1.1011)	(0.1652)	(0.4025)	(0.5842)	(1.520)
Mean tenure² (in years)	-0.405	0.0521	---	---	-0.131	0.0426	0.0012	-0.0445	---	---	-0.0002	-0.0761
	(0.260)	(0.0336)			(0.379)	(0.0289)	(0.0148)	(0.0401)			(0.0218)	(0.0553)
Unit size (in persons)	0.186**	-0.0174	0.160***	-0.0322	0.136*	-0.0264	---	-0.122	0.0464	0.0595	0.0611	0.0647
	(0.0624)	(0.0541)	(0.0491)	(0.0521)	(0.0698)	(0.0502)		(0.148)	(0.06004)	(0.168)	(0.058)	(0.158)
Influenza	0.0602		0.101		0.144		-0.328		-0.5543		-0.816*	
(0=no activity,…)	(0.275)		(0.255)		(0.248)		(0.291)		(0.454)		(0.460)	
Employment protection	6.857		6.423		---		1.924		0.867		---	
legislation (0=weak,…)	(3.882)		(3.743)				(3.875)		(8.749)			
Employment in	106.2		88.02		---		128.565		153.555		---	
manufacturing (in %/100)	(180.0)		(202.2)				(91.666)		(1.981)			
Unemployment rate	-11.46		-30.00		---		3.3119		13.334*		---	
(in %/100)	(96.23)		(96.09)				(14.264)		(26.688)			
Prosperity level (annual	10.56*		10.06*		---		6.1809**		3.495		---	
change in GDP in %/100)	(5.490)		(5.673)				(2.541)		(4.9687)			
2-Shift system (yes=1)	1.442		1.281		0.826		1.921		2.697		1.608	
	(1.083)		(1.136)		(1.229)		(1.262)		(1.699)		(1.577)	
3-Shift system (yes=1)	n/a		n/a		n/a		1.208		1.787		1.261	
							(1.152)		(1.303)		(1.266)	
Constant	3.501		-13.69		50.39		-21.86		-8.137		-16.24	
	(60.35)		(35.07)		(60.45)		(25.73)		(24.844)		(35.19)	
Observations	471		471		471		4,878		2,128		2,128	
R-squared	0.189		0.182		0.157		0.042		0.07		0.068	
Number of units	14		14		14		146		146		146	

Robust standard errors in parentheses | *** p<0.01, ** p<0.05, * p<0.1

+ all estimations including month and weather dummies & age and tenure coefficients of variation | only main results presented | n/a= not available

Table 3.4: Fixed-Effects Interaction Terms on Absence Spell Duration (II)

	luxury sites						**volume sites**					
	II-(1)		**II-(2)**		**II-(3)**		**II-(4)**		**II-(5)**		**II-(6)**	
VARIABLES	**Plant A (GER)**	**Plant B (UK)**	**Plant A (GER)**	**Plant B (UK)**	**Plant A (GER)**	**Plant B (UK)**	**Plant C (ESP)**	**Plant D (GER)**	**Plant C (ESP)**	**Plant D (GER)**	**Plant C (ESP)**	**Plant D (GER)**
Mean age (in years)	-0.276	-5.748	-0.274	0.0202	-0.540	-6.717	-2.24	0.950	0.0628	0.653**	2.566	-1.882
	(2.570)	(4.142)	(0.388)	(0.384)	(2.407)	(4.601)	(2.345)	(2.588)	(0.2221)	(0.302)	(3.603)	(3.091)
Mean age² (in years)	0.000915	0.0673	---	---	0.00411	0.0837	0.0301	-0.00550	---	---	-0.0347	0.0307
	(0.0355)	(0.0470)			(0.0337)	(0.0527)	(0.0317)	(0.0317)			(0.0484)	(0.0379)
Share of males	-21.42*	4.607	-24.3**	6.889	-22.85*	4.248	-2.721	-4.406	-5.2565*	-4.135	-5.017*	-5.513
(in %/100)	(11.35)	(6.781)	(9.569)	(7.457)	(12.78)	(6.197)	(1.789)	(6.571)	(3.0911)	(8.806)	(3.027)	(8.841)
Mean tenure (in years)	-5.607	-0.563	-1.594	0.102	-5.164	-1.798	-0.1032	-0.604	0.4505**	-0.734	-0.0047	1.354
	(4.232)	(1.034)	(1.612)	(0.222)	(4.623)	(1.029)	(0.4351)	(1.583)	(0.2127)	(0.566)	(0.722)	(2.377)
Mean tenure² (in years)	0.295	0.0179	---	---	0.283	0.0462	0.013	0.000880	---	---	0.0167	-0.0799
	(0.278)	(0.0285)			(0.292)	(0.0287)	(0.0174)	(0.0594)			(0.0291)	(0.0926)
Unit size (in persons)	0.0475	-0.0163	0.0735*	-0.0612	0.0691*	0.00821	---	-0.0781	0.2712***	0.399*	0.299***	0.428*
	(0.0362)	(0.0888)	(0.0399)	(0.0704)	(0.0373)	(0.0927)		(0.146)	(0.0845)	(0.236)	(0.0833)	(0.232)
Share of temporary	-0.306		-0.112		-1.233		---		---		---	
workers (in %/100)	(1.902)		(1.704)		(2.073)							
Employment protection	-4.271		-4.687		---		4.748		1.586		---	
legislation (0=weak,…)	(3.690)		(3.655)				(5.002)		(10.27)			
Unemployment rate	-124.6		-133.0		---		9.535		-23.96		---	
(in %/100)	(105.3)		(102.0)				(18.03)		(29.68)			
Prosperity level (annual	1.868		3.022		---		7.583**		2.167		---	
change in GDP in %/100)	(4.328)		(4.416)				(3.558)		(6.486)			
2-Shift system (yes=1)	0.321		0.460		1.037		3.652**		3.606*		3.606*	
	(0.765)		(0.695)		(0.764)		(1.714)		(2.144)		(2.144)	
3-Shift system (yes=1)	n/a		n/a		n/a		2.503*		2.593*		2.593*	
							(1.371)		(1.499)		(1.499)	
Constant	99.98		14.80		120.3*		-4.780		-12.28		-15.59	
	(63.79)		(38.29)		(58.64)		(37.53)		(30.78)		(45.73)	
Observations	471		471		471		4,817		2,129		2,129	
R-squared	0.141		0.134		0.123		0.033		0.052		0.049	
Number of units	14		14		14		143		143		143	

Robust standard errors in parentheses | *** $p<0.01$, ** $p<0.05$, * $p<0.1$

[+] all estimations including month and weather dummies & age and tenure coefficients of variation | only main results presented | n/a= not available

Table 3.5: Fixed-Effects Interaction Terms on Absence Frequency (III)

	luxury sites						volume sites					
	III-(1)		**III-(2)**		**III-(3)**		**III-(4)**		**III-(5)**		**III-(6)**	
VARIABLES	**Plant A (GER)**	**Plant B (UK)**	**Plant A (GER)**	**Plant B (UK)**	**Plant A (GER)**	**Plant B (UK)**	**Plant C (ESP)**	**Plant D (GER)**	**Plant C (ESP)**	**Plant D (GER)**	**Plant C (ESP)**	**Plant D (GER)**
Mean age (in years)	-0.191	0.202*	0.00238	0.0199**	-0.176	0.180	-0.00762	0.0188	0.0024	-1.10e-06	0.0237	0.0461
	(0.188)	(0.110)	(0.0118)	(0.00886)	(0.193)	(0.109)	(0.0244)	(0.04497)	(0.0027)	(0.00428)	(0.0384)	(0.0590)
Mean age² (in years)	0.00261	-0.00205	---	---	0.00238	-0.00198	0.00011	-0.00025	---	---	-0.00028	-0.00056
	(0.00267)	(0.00121)			(0.00273)	(0.00122)	(0.00032)	(0.00054)			(0.0005)	(0.00069)
Share of males (in %/100)	1.074	0.0559	1.206*	0.0700	1.152	-0.0761	-0.056**	0.0593	-0.087**	0.04906	-0.081**	0.0201
	(0.626)	(0.301)	(0.598)	(0.327)	(0.653)	(0.293)	(0.0256)	(0.0795)	(0.0368)	(0.113)	(0.0362)	(0.107)
Mean tenure (in years)	0.402**	-0.0723	-0.0490	-0.0112*	0.293*	-0.0609	0.0044	0.0442	0.0018	-0.0002	0.0072	0.0370
	(0.157)	(0.0612)	(0.0517)	(0.00594)	(0.165)	(0.0577)	(0.0055)	(0.0371)	(0.00094)	(0.00781)	(0.00797)	(0.0419)
Mean tenure² (in years)	-0.0331**	0.00173	---	---	-0.0249**	0.00168	-0.0002	-0.00139	---	---	-0.00027	-0.00141
	(0.0115)	(0.00174)			(0.0114)	(0.00166)	(0.00021)	(0.00139)			(0.00031)	(0.00152)
Unit size (in persons)	0.00392*	-0.00297	0.00211	-0.00176	0.00263	-0.00304	---	-0.00105	0.00104	-0.00228	0.00104	-0.00233
	(0.00220)	(0.00176)	(0.00190)	(0.00138)	(0.00234)	(0.00175)		(0.0033)	(0.00094)	(0.00340)	(0.00094)	(0.00335)
Turnover	0.000184		0.000300		0.000199		0.0011*		0.00157**		0.00179**	
(arrivals & exits)	(0.000365)		(0.000412)		(0.000365)		(0.00064)		(0.00082)		(0.000814)	
Share of temporary	0.105		0.0847		0.157*		---		-0.0114		-0.00426	
workers (in %/100)	(0.0768)		(0.0920)		(0.0815)				(0.0598)		(0.0565)	
Employment protection	0.161**		0.159**		---		-0.0984		-0.0122		---	
legislation (0=weak,…)	(0.0638)		(0.0680)				(0.0782)		(0.1597)			
Employment in	6.746		4.146		---		3.081		5.511**		---	
manufacturing (in %/100)	(5.214)		(4.832)				(1.909)		(3.0594)			
Unemployment rate	-0.359		-0.905		---		-0.0436*		0.6342		---	
(in %/100)	(1.980)		(2.365)				(0.2449)		(0.641)			
Prosperity level (annual	0.245		0.179		---		0.000921		0.3009***		---	
change in GDP in %/100)	(0.156)		(0.172)				(0.000630)		(0.1112)			
3-Shift system (yes=1)	n/a		n/a		n/a		0.0201		0.0290*		0.0268	
							(0.0155)		(0.0174)		(0.0173)	
Constant	-3.049		-1.400		-1.242		-0.1187		-0.8429*		-0.595	
	(1.977)		(1.045)		(2.060)		(0.4345)		(0.5814)		(0.611)	
Observations	471		471		471		4,810		2,123		2,123	
R-squared	0.272		0.239		0.250		0.053		0.064		0.061	
Number of unit	14		14		14		143		143		143	

Robust standard errors in parentheses | *** p<0.01, ** p<0.05, * p<0.1

+ all estimations including month and weather dummies & age and tenure coefficients of variation | only main results presented | n/a= not available

In H2, we have hypothesized that absence usually increases with unit size due to better shirking opportunities in larger teams (e.g. free riding is facilitated as managerial control is complicated and team members are less familiar). Concerning absence rate, we find support for the predicted relationship only at plant A. Here, a growth in unit size by one person increases absence rate by 0.186 percentage points (significant at the 5%-level) in our preferred model specification (I-1, -2, -3). This relationship is confirmed for absence spell duration at sites A (II-2, -3), C (V-3) and D (V-4). While the findings for plant A display only moderate increases of absence spell duration between 0.069 to 0.074 days with each additional employee, findings for plant C (0.271 to 0.311) and D (0.399 to 0.428) suggest a steeper increase. However, we do not attribute these results to shirking behavior since absence spell duration is understood as a measure of involuntary absence. Instead, we assume attendance management by shop floor supervisors to be more difficult in larger teams (e.g. less time to care for individual needs and work-related and non-work-related health issues). The positive impact of a communicative and trustful personal manager-employee relationship on employee health has been emphasized by almost all on-site experts. They confirmed that in small teams, managers are more familiar with health issues and personal circumstances of their subordinates and, therefore, can more easily take preventive actions. Recent research confirmed the beneficial effects of high-quality supervisor-employee relationships on employee absence (e.g. Tenhiälä et al. 2013). Eventually, we observe inconsistent findings for the relationship between unit size and absence frequency (positive at plant A (III-1), negative at plant B (VI-2)). Still, we find support for hypothesis H2, yet for different reasons than initially predicted.

Concerning the share of temporary agents in a unit, we find mixed results with regard to absence rate.[26] On the one hand, we observe a positive link at site B (IV-2) since absence rates increase with the share of temporary workers. On the other hand, we observe the opposite effect to be true at plant A (IV-1) with absence rates decreasing as the share of temporary workers increases. Both effects seem to be reasonable to some degree when we assume temporary agents to be incentivized to refrain from absence due to their contract status (e.g. Engellandt, Riphahn 2005, Bradley, Green, Leeves 2007). This assumption is backed by on-site experts who report that temporary agents have nearly zero absence due to two main reasons. First, temporary agents do not benefit from sick pay regulations to the

[26] The share of temporary workers has only been reported by plants A, B and D. Thus, we had to exclude this variable from interaction term estimation of plants C and D.

same extent than do permanent employees. Second, absence is usually one of the most important criteria when deciding on employing temporary agents permanently. Taking this into account, two alternative interpretations might explain the mixed results. On the one hand, absence might decrease with the share of temporary agents as permanent employees are motivated to display low absence by means of productivity spillovers (e.g. Mas, Moretti 2009). On the other hand, absence may increase since permanent employees intensify shirking. The latter argument is supported by findings on voluntary absence as proxied by absence frequency. We observe a positive link between the share of temporary agents and unit-level absence of permanent employees at sites A and B (III-3, VI-1). Thus, in units with a large share of temporary agents permanent workers seem to increase shirking. Yet, given the mixed results on absence rate, we can confirm hypothesis H3 only for voluntary absence at sites A and B.

Data on the share of team members with health impairments has only been provided by both volume car producers C and D. Yet, in none of our model specifications we observe a significant effect on overall team absence. In other words, although workers with health issues are more prone to absence and might not be able to perform all tasks in the work area of the team, their impairments do not affect overall unit absence. We assume organizational activities such as job rotation to support workers suffering from constrained health. Hypothesis H4 could not be confirmed.

Economic influences / incentives (ECON)

We consider four potential economic determinants in our analysis: sickness benefits, employment protection legislation, (un)employment situation and prosperity level. In order to check for influences of divergent national sickness benefits, we re-estimated our models using simple OLS regression models since fixed effects models treat the variable as plant fixed effect due to missing variation in national legislation throughout the observation period.[27] To our surprise we are not able to find any statistically significant relationship between sickness benefits and neither absence rate, nor duration or frequency. This is even more surprising since existing research comprehensively confirms the influence of national sickness benefit regulations on employee absence behavior (e.g. Frick, Malo 2008, Johansson, Palme 2002, 2005). However, we learned from on-site experts that both international producers supplement statutory sickness benefits with voluntary additional sick pay for

[27] Since results did reach statistical significance, OLS and random-effect estimations are not reported in this paper due to reasons of brevity.

their employees. In other words, national differences in statutory sickness benefits are leveraged by the employer and, therefore, are of no importance in the organizational setting studied in this paper. Therefore, we refrain from empirically testing hypothesis H5 based on our data.

In hypothesis H6, we assume a strict employment protection legislation to increase employee absence behavior because punishing or even dismissing workers for absenteeism should be complicated due to protective employment rights. Again, we re-estimate our models with random effects specifications to account for the fact that employment rights only changed in Spain and the UK over the three-year observation window. However, results did not change. Absence rate and absence spell duration increase only at the UK site with stricter employment protection laws (IV-2). Moreover, absence frequency can be observed to increase in pooled estimations at plants A and B (III-1, -2, VI-2), probably due to the fact that differences in employment protection laws between Germany and the UK are more pronounced than between Germany and Spain. In total, we find only weak empirical support for the relationship predicted in H6.

Concerning (un)employment, we observe significant effects at both volume car producers. An increase in the share of employment in manufacturing induces an increase in absence frequency at plants C and D (III-5). Referring to absence frequency as proxy for voluntary absence it seems as if employees increase shirking when they assume to have good outside job opportunities. This finding is confirmed by a significant reduction of absence frequency when unemployment figures increase (III-5). Again, employees tend to increase shirking when overall employment is high and they do not have to fear long spells of unemployment. This is in line with hypothesis H7. However, we observe absence rates to increase with unemployment in one model specification at plants C and D (I-5). Thus, we cannot fully support hypothesis H7 due to inconclusive findings.

Eventually, we explore hypothesis H8 predicting that with increasing prosperity level absence may be more pronounced. We found support for this argument, since pooled estimations for plants A and B (I-1, -2) as well as C and D (I-4) suggest absence rates to increase with positive annual changes in GDP. The same is true for absence duration and absence frequency at plants C and D (II-4, III-5). In other words, employees increase their absence during a nationwide positive economic situation while decreasing absence in hard economic times. We believe the interpretation to be straightforward. Employees tend to shirk in an overall positive economic environment as they do not have to fear job-loss or believe to have attractive outside options. This behavior has been observed by on-site experts at the Spanish site throughout the financial crisis of the last years. With the beginning of the economic downturn, absence decreased dramatically while it starts to rise again due to the recent economic recovery. All in all, we are able to confirm hypothesis H8.

Worker characteristics (WORK)

We address potential influences of worker characteristics on absenteeism along the four dimensions of gender, age, tenure and acute health condition. Concerning gender, we observe a significant relationship between absence behavior and the gender composition of a work unit at plant C. Here, absence rates are lower in all-male vs. all-female teams. In other words, a higher share of females in a team is detrimental to work group absence. This result is in line with existing research on gender differences in absence behavior (e.g. Barmby, Ercolani, Treble 2002). The gender-absence relationship is often assumed to stem from higher daily burdens of job and family life for females since child (or elder) care is still more associated with women than men (e.g. Åkerlind et al. 1996, Bratberg, Dahl, Risa 2002, Vistnes 1997). On-site experts at plant C confirmed this reasoning as the role model of the mother being responsible for child care is deeply entrenched in the Spanish culture. For absence spell duration we find a similar pattern – again only valid at plant C (II-5, -6).[28] Keeping in mind that absence duration is often classified as involuntary, one might assume that women suffer more from long-term illnesses which arise from burdening working conditions than their male colleagues. However, units with a larger share of males do not only display shorter absence spell durations at plant C, but also lower absence frequency (III-4, -5, -6, VI-3).[29] Two potential explanations seem reasonable. First, research often links female absence with short-term reproduction-related health issues such as menstruation (e.g. Alexanderson et al. 1996, Sydsjö, Sydsjö, Alexanderson 2001). Second, acknowledging that women are more involved in child or family care than men one may assume that female workers might stay at home without being ill in order to nurse their sick child. Both interpretations are confirmed by on-site experts and may explain the larger frequency of absence coming along with a higher share of female workers. Overall, there seems to be a statistically significant relationship between the gender composition of the

---

28 Although we find similar results for the German luxury car producer (A), we refrain from including these findings into our considerations since variation in gender composition is extremely low at this plant (mean share of males exceeds 97%).

29 Again, a significant finding for plant A is not further considered due to the high average share of males.

unit and absence behavior – at least at plant C. Here, units with a larger share of women show significantly higher absence rates, longer absence spell durations and higher absence frequencies. Therefore, we are able to confirm H9 for the Spanish volume site.

We predict two potential effects of unit mean age on absence. On the one hand, H10 states that involuntary absence increases with age as workplace burdens are getting harder to cope with when getting older. On the other hand, we expect voluntary absence to decrease with age as older workers face higher financial and social obligations than younger workers (H11).[30] Findings on absence rate reveal a positive relationship with age at plant D (I-5, IV-4). With each additional year of unit mean age, absence rates increase by 0.501 to 0.516 percentage points at this plant. The same result holds true at plant B (I-2). Here, a one year increase is associated with a raise of 0.7 percentage points. These effects might seem astonishingly high at first sight, but one has to take into account that workforces are quite homogeneous concerning age at all four sites (standard deviation of units' mean age is 3.45 at max). We further tested for non-linear age effects by including squared age, but in none of our model specifications and for none of the dependent variables the squared term of age reaches statistical significance. Therefore, we assume the link between age and absence rate to follow a linear relationship. The very same relationship can be found at plant D for absence spell duration – the classical measure of involuntary absence (II-5, V-4). With a one year increase in a unit's mean age, absence spell duration increases between 0.653 to 0.678 days per month. We, therefore, are confident to confirm hypothesis H10 for plant D. We assume this effect to arise from workplace burdens that become substantially more difficult to meet with increasing age. On-site experts confirm our reasoning. Fairly unexpected, we observe a positive relationship of age and absence frequency in our preferred model specification of plant B (III-2). Here, a one year increase of unit mean age yields at an increase of absence frequency by 0.02 points. We do not find any other statistically significant effect with regard to absence frequency. This is even more astonishing since on-site experts have confirmed our hypotheses on voluntary absence in reporting older employees to have a more rigorous zero-absence tolerance while younger generations seem to be socialized with a less serious attitude towards absence culture at work. Still, based on the empirical findings we have to reject hypothesis H11.

[30] In order to address concerns that tenure and age are usually highly correlated (Barmby, Ercolani,Treble 2002, Gordon, Johnson 1982), we re-estimate all models excluding tenure. Leaving out tenure variables, however, does not change any results and, in particular, does not influence our findings for age.

When it comes to tenure, we have predicted in H12 that absence decreases with tenure since more experienced workers should have developed better coping strategies over time which help them to bear workplace risks and burdens. With absence rate being the dependent variable we observe the expected negative relationship at plants A (I-2, IV-1) and B (I-1, IV-2). Likewise to age, the high effects may arise from low overall heterogeneity within the workforce regarding tenure (unit-level standard deviation of mean tenure reaches 3.31 at max). Throughout discussions with on-site experts we learned that experience is a major asset helping older workers to outperform younger ones on many occasions. According to the experts, being “an old hand” often means knowing best how to tackle tasks and problems and to be used to workplace burdens from experience. We included squared terms to control for non-linear effects of tenure. Results suggest a u-shaped relationship between absence rate and tenure at plant B (IV-2). In other words, the predominantly negative relationship between absence rate and tenure is getting convex to a certain degree with increasing tenure. With regard to absence spell duration, we find a positive link with tenure at plant C (II-5). This result is somehow surprising as it – against our predictions – suggests absence spell duration to increase with tenure. Squared tenure terms do not reach statistical significance indicating linear relationships between tenure and absence spell duration. Concerning absence frequency, we observe inconsistent findings in terms of an inverse u-shaped relationship at plant A (III-1, -3, VI-1) and a negative linear relationship at plant B (II-2). In total, there is some support for our predicted relationship since absence rates decrease with tenure at both luxury car production sites. In contrast to this, absence spell duration increases with tenure at plant C. Yet, results for absence frequency neither reject nor support our hypothesis H12 due to inconsistent findings. Thus, overall results on H12 remain inconclusive.

Interestingly, we observe only a very weak relationship between employee absence and acute illnesses as proxied by national influenza activity. In particular, only two model specifications display significant but opposing effects, providing no clear-cut evidence on ILI determining absence (I-6, V-3). This is even more astonishing as literature confirms a strong relationship (e.g. Tsai, Zhou, Kim 2014, Keech, Scott, Ryan 1998) and on-site experts report influenza outbreaks to be clearly detectable in their daily absence data. We believe our national influenza measure to somehow dilute the findings since it does not allow for a further differentiation in local outbreak intensity. All in all, we could not provide support for hypothesis H13 – there is no link between absence and acute illness measured by influenza statistics detectable in our data.

Working condition variables (COND)

Following empirical results by Slany et al. (2014), Morikawa et al. (2001) and Fekedulegn et al. (2013), we predict in H14 that working hours other than the usually applied one-shift systems (day) increase absence due to negative health and social impacts coming along with irregular and/or atypical working hours. In our estimations, we observe two-shift systems to increase absence rates in comparison to the reference one-shift system at both international plants B and C by 2.8 (IV-2) and 2.3 (IV-3) percentage points. For plant A, we find a significant effect in the opposed direction (IV-1). However, we believe this result to be highly biased since at plant A only three units worked a one-shift system for only six months, while operating on a two-shift system for the remaining observation period. Concerning absence spell duration, pooled estimations for both volume car producers C and D reveal statistically significant effects of two-shift and three-shift systems (II-4, -5, -6). Plant-level estimations for two-shift systems at plants B (V-2) and C (V-3) and for a three-shift system at plant C (V-3) support this finding. Referring to absence frequency, we find a weak support for the negative impact of a two-shift system at plants B (VI-2) and C (VI-3) and a three-shift system at plants C and D (III-5). On-site experts acknowledge that night shifts are usually the most exhausting working hours for the majority of the work force as they are opposing the human sleeping pattern. In particular, the transition from night shift to the subsequent day shift is exceptionally burdening. Interestingly, experts report that absence is lowest in night shift weeks since employees do not want to forfeit night shift premiums which can amount up to 30% of their normal pay (see Stein (2015) for empirical support). Still, we are able to partially confirm our hypothesis H14. Additional shifts significantly increase employee absence at some plants.

Control variables (CONT)

Although no hypotheses were derived for date and weather controls, our estimations reveal some interesting insights.[31] At both volume car producers, absence rates and spell duration are significantly lower during most summer months from April to September in comparison to the reference period (January) – even when influenza activity and weather is controlled for. However, this phenomenon is most likely attributable to the summer holidays

[31] Results for control variables are not tabulated in this paper.

as absence is usually not reported while being on holiday. Findings for absence frequency are weak and inconsistent with no clear pattern. Concerning weather, we find support for arguments by Shi and Skuterud (2015) who state that employees adapt their absence behavior to the weather conditions. At the Spanish site, we find absence rate, duration and frequency all increase significantly with temperature. For instance, unit-level absence spell duration increases by 0.47 days with each degree Celsius at the 1%-level of significance. This finding is in line with the impressions of on-site experts at this plant. Same results are found at least for absence spell duration at plant D (0.2 days, 1%-level). No other effects of temperature, sunshine hours and precipitation are found. In other words, we observe a link between weather and absence at some plants indicating that employees may adapt their absence behavior to favorable weather conditions. This behavior can be interpreted as shirking.

In order to control for the robustness of our findings, we re-estimate our models using quantile regression.[32] Although this approach delivers additional significance at some quantiles, overall results remain the same. Furthermore, we pool our data over all four plants. Yet, we fail to find overall statistical patterns. Eventually, we limit our data to units working in the assembly line since assembly line production is the only division that is observed in all four plants. Yet, this does not deliver further insights on potential determinants of worker absenteeism.

Unfortunately, our study entails limitations at some instances. First, by focusing on aggregated data we gain evidence on team-level determinants of absence, yet at the expense of limited insights on individual-level influences. Second, although using data on absence spell duration and frequency to proxy for voluntary and involuntary absence, our study is limited in so far that both measures may contain their respective counterpart as well (Thomson, Griffiths, Davidson 2000). Third, we are not able to control for individual or group-level health conditions of employees other than by applying nation-wide WHO influenza outbreak statistics. In doing so, we are quite confident to at least proxy for employee health conditions concerning acute flu pandemics. However, we do not have an appropriate measure for other illnesses such as musculoskeletal disorders or stress-induced illnesses provoked by monotony, repetitiveness and cycle time. Yet, we assume physical as well as mental sicknesses to be distributed homogeneously among shop-floor teams after

[32] Results are not reported in this paper.

controlling for age and gender. Fourth, by working on monthly intervals we only provide evidence on determinants that are relatively stable over time (Harrison, Martocchino 1998). Conversely, this means that daily or other short-term effects on absenteeism are leveled out over the aggregated monthly periods and, therefore, could not be studied in this paper.[33] Fifth, we are not able to control for educational and cultural influences on absence behavior. This would have been particularly interesting since discussions with on-site experts have shown that absence cultures vary widely between cultural spheres. Additionally, we are not able to include more detailed workplace characteristics other than the implemented shift system to assure the anonymity of teams. Eventually, since this paper is built on insider data, findings might suffer from a lack of generalizability to other working conditions and organizations.

3.6 Conclusion and Implications

Using a hitherto unavailable data set gathered from insider econometric data of four international manufacturing plants of a large European automobile company, we are able to provide evidence on determinants of absenteeism. In total, we investigate team-level absence of blue-collar employees organized in 160 work units (approx. 2,150 workers). Based on a comprehensive literature review we derive fourteen hypotheses covering four main areas of potential determinants on absence: social peer influences, economic influences / incentives, worker characteristics and working conditions. We focus our analysis on three unit-level dimensions of absence: absence rate (percentage of scheduled time lost due to absence), absence spell duration (average number of days lost per absence occasion) and absence frequency (number of absence incidents divided by unit size). Our empirical analyses follow a two-stage research strategy that complements insider econometric data analyses with interviews with on-site experts. Interestingly, we observe effects to vary by production site, however, with only very few consistent findings. Since a detailed description of results by plant has been given throughout chapter 3.5, we refrain from recapitulating effects by site in what follows, but instead summarize the overall effects.

Considering social level determinants, we predicted turnover to decrease absence since group norms are less enforceable when team composition is changing. Yet, results are

[33] Although empirical evidence is missing at this point, we were able to gain some anecdotal insights by means of the experts' discussions. For instance, experts confirmed findings by Rosenblatt, Shapira-Lishchinsky and Shirom (2010) of increased absence around school holidays and weekends ("Monday morning flu"). At one plant, experts could also identify higher absence rates at match days of the local soccer team in the affected shift and, thus, support arguments by Thoursie (2004).

pointing in the opposite direction, presumably due to a feeling of uncertainty and the lack of an overall team spirit in teams with high turnover. In accordance with our expectations, we observe absence to increase with unit size. Although results may suggest shirking, we follow on-site experts in arguing that line managers' attendance management may be more caring and familiar in smaller teams. With regard to peer effects, the share of temporary team mates is positively related to voluntary absence of permanent workers, presumably due to shirking. Surprisingly, the share of workers with health impairments fails to reach statistical significance in all model specifications. Evidence on economic influences and incentives reveals the expected positive link between employment protection laws and absence. In other words, workers tend to increase their absence behavior when feeling protected by strict employee rights. Moreover, we hypothesized a negative relationship between absence and outside options expressed by national (un)employment figures. Surprisingly, we observe this relation to be inconclusive. Furthermore, we are able to confirm the expected positive link between prosperity level and absence behavior. This effect seems to be attributable to a certain feeling of job security since on-site experts observed the inverse relationship during the financial crises of the previous years. We were not able to properly study the effects of national sickness benefit systems on absence since the company at hand grants additional sick pay to compensate for weaknesses in statutory sickness benefits. As a consequence, differences in national legislation are leveled out. Regarding worker characteristics, results suggest that absenteeism increases with the share of female employees within a unit. According to on-site experts, this effect is to a large degree attributable to the double burden of work and family obligations which is even today substantially more pronounced for females than for males. For some sites, we were able to confirm the predicted positive relationship between age and absence rate and spell duration. Thus, it seems that meeting workplace burdens is substantially more demanding with age. Yet, against our expectations we observe weak evidence that voluntary absence, too, increases with age. Results for mean organizational tenure and acute health conditions expressed by national influenza outbreak statistics are inconclusive. Consistent with the hypothesized relationship, we find two-shift as well as three-shift systems to induce significantly higher absence than the standard one-shift system. According to on-site experts this effect is largely attributable to the non-correspondence of working hours and human bio-rhythm as well as social life. All findings are robust on various specifications and survive a large number of robustness checks.

Subsequent to the data analyses, we conduct a total of 18 expert interviews at both international plants with shop floor staff, worker representatives, line supervisors and HR managers in order to double-check our own experience from the national plants. These interviews helped us to further interpret our results and to learn more about the instruments of attendance management in practice at other locations. Based on our empirical findings and the practical experience of on-site experts, we yield some important implications that may help to increase both overall health conditions and attendance of blue-collar workers.

The results on unit size seem to suggest that small units are advantageous over large units in terms of employee attendance. As discussed, we predicted this phenomenon to arise from better shirking options in larger teams. However, we learned from experts that quality of and familiarity within the manager-employee relationship is moderated by team size. We, therefore, suggest managers to invest in developing trustful and sincere social ties with their subordinates based on a fair and open culture. Of course, building up intensive relations within teams requires time line managers usually do not have. Thus, we call for HR and plant management to support line managers by granting a sufficient amount of time to invest in socializing with their teams. We believe the long-term outcomes in terms of reduced absenteeism will leverage any initial time loss.

With regard to economic influences it seems to be infeasible for a company to take actions since politics (e.g. sickness benefits, employment protection) as well as overall economic situation (e.g. unemployment, prosperity level) are usually out of the range of influence of a company. Yet, it might be a reasonable strategy to openly communicate the current economic situation of the company, in particular the competitiveness on the global markets, to the employees to increase their awareness of potential menaces. This could motivate employees who feel too secure. As mentioned above, both international plants compensate for weaknesses in statutory sick pay by offering additional sickness benefits. However, at one site employees with more than three non-work related absence occasions per 12 months are getting an official warning and are withheld additional sickness benefits on their next absence incident (repeated absence arising from known long-term sicknesses e.g. cancer is not taken into account). According to on-site experts, this policy has proven to be successful in reducing absence frequency. Overall, empirical evidence on environmental determinants is pointing in the direction that drawing international comparisons of absence figures within the company might be biased due to national variance in laws and economic situations. In other words, critical assessments of international locations are not necessarily accurate and true and, as a consequence, policy decisions based on international comparisons may be misleading.

In response to empirical evidence confirming the predicted positive link between age and absence, we emphasize the importance of harnessing all possible health-promoting potentials in ergonomic workplace design and occupational health care (OHC). We acknowledge that at all four plants working conditions are constantly improving in terms of ergonomics. Additionally, it seems important to mention that medical attendance of musculoskeletal illnesses by OHC is very comprehensive and high in quality. The same is true for medical check-ups offered to employees for free in the company's on-site OHC centers. However, we learned from local experts that work-related and non-work related mental stressors are of increasing importance even in manufacturing. Experts reveal that on-site OHC centers still lag in developing a strategy to treat stress-induced illnesses as comprehensively as musculoskeletal illnesses. Moreover, some findings emphasize the advantageousness of teams that are mixed with regard to age since younger worker can benefit from older ones and vice versa. It, therefore, seems reasonable to yield a heterogeneous team composition.

Considering empirical evidence suggesting females to suffer from the job-family double burden more than men it seems important to support families in dealing with the reconciliation of work and family life. From on-site experts we learned that a different set of supporting activities is in practice at each site. Instruments that enhance individual flexibility have proven particularly helpful in increasing attendance of parents. We, therefore, suggest offering parents high flexibility given organizational constrains. For instance, at one location parents can work reduced hours even when working shifts. Here, two persons on reduced shifts share one job. In situations where both parents are employed on the shop floor, they are offered opposing shifts if necessary to guarantee child care around the clock. At some plants, employees can use dependency leave (or third-party leave) when children or elders are ill and need intensive care. Although third-party leave has proven to be very useful in allowing parents to care for their sick children without having to fake own illnesses, employees report that making use of this instrument is often criticized by supervisors and HR. Greater understanding seems to be necessary at this point.

Eventually, our findings on shift systems prompt the thought that shift-work should be avoided as we observe two-shift and three-shift systems to increase absenteeism in comparison to the normal one-shift (day) systems. Yet, from local experts we learn that it is sometimes preferable to add a second or third shift to better distribute workload instead of using overtime. If shift work is unavoidable, it seems important to follow the latest scientific insights on healthy shift work design.

In general, the findings of this work depict an important contribution to (personnel) economics absence research. In particular, we benefit from exceptional and unique insider data that allow for comprehensive econometric analyses. However, in this study we only observe automobile workers. As a consequence, we might suffer from selection bias that entails endogeneity problems – yet a common issue in insider econometric studies. We, therefore, encourage other researchers to further study absenteeism in different organizational contexts in order to gain a comprehensive idea of the functioning of employee absenteeism and to refine ways to manage attendance.

4 On the Road Again: Crowding-Out Effects of Extrinsic Motivation in Commercial Trucking

4.1 Introduction

Today, the use of variable compensation schemes in the form of exogenous financial incentives has become a common practice in organizations to align interests of principals and agents.[34] In this paper, we seek to study how extrinsic rewards may influence employee effort choice decisions and, thus, worker productivity. We analyze performance under varying incentive schemes based on unique data on commercial truck drivers. The trucking industry appears particularly suitable for this purpose since the carrier-driver relationship represents the classic principal-agent setting as modeled by economic theory (e.g. Oyer, Schaefer 2011) with agents producing their output far beyond the monitoring range of the principal.[35] Additionally, truck driver performance is entirely achieved on an individual level and, therefore, is not prone to any kind of team bias.[36] To solve this very specific principal-agent problem, asset ownership is often considered a particularly cheap and efficient monitoring instrument. Yet, its applicability is limited since truck drivers are assumed to be risk averse and limited in their access to capital (e.g. Arrunada, Gonzalez-Diaz, Fernandez 2004, Nickerson, Silverman 2003, Sheikh 2007). Recently, modern GPS-based fleet management systems advance driver monitoring from only observing the simple output (e.g. arrival time and goods carried) to more comprehensive real-time surveillance of driver performance while on the road (e.g. Baker, Hubbard 2004, Barla et al. 2010, Hubbard 2000). This allows for new approaches in incentive design for drivers.

Trucking is, by and large, an increasingly competitive business environment for haulers as well as drivers. Nowadays, fuel costs account for around 30% of the total life-cycle costs of a heavy duty truck (e.g. Schittler 2003). In this situation, findings that only slightly increase operational efficiency may leverage profitability and, therefore, are appreciated by practitioners. At the same time, the road transportation industry is a substantial consumer

[34] Exogenous material incentives as understood in this paper to comprise any wage configurations that at least partially involve variable remuneration such as bonus payments, piece rates and pay for performance schemes. Although not existent in the organizational setting at hand, there might be non-monetary extrinsic incentives as well (see, for instance, seminal papers on tournaments by Rosen (1986) and career concerns by Fama (1980)).

[35] Vernon and Meier (2012) discuss the peculiarities of principal-agent settings in trucking in more detail.

[36] Still, exogenous influences determine driving behavior to a large extent, e.g. weather, road conditions and traffic density. Research that does not include these confounding impacts most likely provides incorrect and fallacious conclusions. Thanks to our extensive data, we are able to control for various external influences. All control variables are discussed in the remainder of this paper.

of fossil fuels. Within the EU, CO_2 emissions of heavy duty vehicles – trucks, buses and coaches – have increased by 36% between 1990 and 2010 and now account for about 25% of overall CO_2 emissions and 5% of all greenhouse gas emissions from road transport (European Commission 2014a). Thus, motivating commercial truck drivers to adapt fuel efficient driving provides a promising opportunity to contribute significantly to the reduction of both fuel consumption and CO_2 emissions for economic and environmental benefits. This is where monetary incentives for drivers come into play.

Academic theory knows two opposing effects of exogenous incentives on motivation. Classic personnel economics theory assumes exogenous incentives to increase employee effort choice and productivity by aligning the interests of principals and agents (e.g. Gibbons, Roberts 2013, Oyer, Schaefer 2011). There is broad empirical and experimental evidence on this issue in varying organizational settings (e.g. Kahn, Silva, Ziliak 2001, Lavy 2002, 2009, Lazear 2000b, Shearer 2004). Nevertheless, behavioral incentive theory assumes exogenous incentives to crowd out intrinsic or social motives that are found to be equally important in determining employee effort choice and productivity (e.g. Bénabou, Tirole 2003, 2006, Gneezy, Meier, Rey-Biel 2011). As a consequence, overall employee performance might be lower when being monetarily incentivized. Again, there is comprehensive support for these arguments in the literature (e.g. Ariely, Bracha, Meier 2009, Frey, Oberholzer-Gee 1997, Gneezy, Rustichini 2000a, 2000b).

In response to the concurrent theoretical assumptions and the conflicting findings on the effects of extrinsic incentives, this study aims at analyzing employee effort choice decisions under varying incentive schemes given the specific organizational setting of the trucking industry. We use hitherto unavailable data on the performance of 37 commercial truck drivers. Data is gathered from the GPS-based fleet management system of an in-house hauler of a large European truck manufacturer. The particular features of this research are twofold. First, in contrast to most studies on extrinsic incentives we do not focus on the introduction of a bonus pay scheme, but on its abolition. Second, our analyses are based on four performance measures recorded by in-vehicle computers of the hauler's fleet management system that collect real-time information on truck positioning, technical parameters and driving behavior. This data allows us to contribute to the ongoing discussion on the benefits and pitfalls of extrinsic incentives. If classic personnel economics theory holds true, we expect drivers to perform worse after the incentives are abolished. In contrast, if behavioral incentive theory holds true, we expect driver performance to increase although drivers are no longer extrinsically incentivized. The empirical evidence of this work reveals a significant increase in driver performance after incentives have been abolished, independent of the performance measure. Thus, our findings confirm assumptions stated by behavioral incentive theory as extrinsic incentives are shown to reduce employee performance. Following previous research (e.g. Frey 1994, Frey, Oberholzer-Gee 1997, Gneezy, Meier, Rey-Biel 2011), we attribute this phenomenon to the crowding-out effect of extrinsic incentives. Other important intrinsic motives – such as internal competition or drivers' environmental believes – are literally bought off by monetary incentives.

With this study, we contribute to the literature in several regards. First, we add to the current discussion on the advantages and drawbacks of extrinsic incentives on worker productivity by extending the empirical literature on employees' effort choice decisions under different incentive schemes. Our findings confirm the crowding-out approach suggested by behavioral incentive theory. Second, while numerous studies analyze the switch from fixed pay to performance pay (e.g. Lazear 2000b, Shearer 2004), little evidence is provided on situations where incentives are completely abolished. One of the few exceptions is a study by Freeman and Kleiner (2005) who study the change from piece rate pay to hourly wages. As a result, they observe a significant productivity loss. We add to their research by offering a comprehensive analysis of employees' behavioral responses to the abolition of extrinsic incentives. Third, to our knowledge we are the first to use truck driver performance data gathered by modern GPS-based fleet management systems to analyze employee behavior. This seems surprising since this data consists of broad objective performance information. We, therefore, hope to pave its way into personnel economics. Finally, our work contributes to the emerging insider econometrics approach introduced in seminal papers by Ichniowski, Shaw and Prennushi (1997) as well as Lazear (2000b). This approach applies elaborate econometric methods to (panel) data of only one organization. Thus, it delivers high-quality findings on a micro-perspective, yet at the expense of a limited generalizability.

This paper proceeds as follows. The next paragraph discusses the two competing theories on the impact of extrinsic incentives on employee behavior. Section three describes the unique organizational setting of this study while section four explores the estimation strategy as well as the results of our econometric analyzes. Eventually, section five offers an interpretation and final concluding remarks.

4.2 Theoretical Framework

Exogenous incentives within organizations may affect employee behavior and worker productivity in two ways. First, extrinsic rewards trigger workers to behave as intended by the employer by means of a direct price effect (e.g. Lazear 2000b, Shearer 2004). Second, at the same time exogenous incentives induce psychological impacts that function against the price effect by crowding out the intended behavioral response (e.g. Frey, 1994, Gneezy, Meier, Rey-Biel 2011). On the basis of these assumptions, two competing theories are discussed. On the one hand, personnel economics theory broadly supports the prize effect argument of increasing effort and productivity if employees are incentivized exogenously. Thus, monetary rewards are assumed to reduce information asymmetries in classic principal-agent settings by aligning the interests of principals and agents (e.g. Oyer, Schaefer 2011). On the other hand, behavioral incentive theory reckons the existence of a crowding-out effect that presumes exogenous incentives to undermine intrinsic motivation and, thus, to decrease employee effort and productivity. Experimental and empirical evidence on both conceptions is broad and convincing over different research settings. In the following, we will discuss both theoretical approaches and offer a brief review of the respective findings discussed in the literature.

Classic personnel economics theory is built on the assumption that human beings respond to exogenous incentives by increasing their individual effort (contract theory).[37] Agency theory is centered on the interaction of a principal who motivates one or more agents to perform a task, e.g. by paying a monetary wage. While the agents bear all costs of effort in terms of individual disutility, the principal entirely benefits from the employees' productivity. Thus, as long as not otherwise motivated, agents will not choose effort levels above those minimum requirements that ensure persistent salary. In this situation, monetary incentives in terms of pay for performance may align the interests of principal and agents by linking, at least partially, compensation to effort. In other words, agents' maximum utility changes with higher effort levels. Still, the principal might not be able to measure employee effort efficiently and, therefore, information asymmetries may arise. Again, exogenous incentives may help to overcome information asymmetries by deterring agents from shirking (for a detailed discussion of incentives in principal-agent settings see, among others,

[37] According to Lazear (2000b) it is a "*cornerstone of the theory in personnel economics that workers respond to incentives*" (p. 1346). Other authors regard monetary incentives as „*the essence of economics*" (Prendergast 1999, p. 7) or a "*central tenet of economics*" (Bénabou, Tirole 2003, p. 489).

Gibbons 1998, Gibbons, Roberts 2013, Oyer, Schaefer 2011). There is broad experimental and empirical evidence that exogenous incentives enhance productivity.

In his seminal paper on windshield installers Lazear (2000b) observes productivity to increase by on average 44% after the company shifted from fixed wages to piece-rate pay. However, only half of this raise seems to be attributable to individual-level productivity gains induced by pay for performance compensation.[38] In a field experiment, Shearer (2004) finds a similar price effect observing a more than 20% productivity increase for tree planters following the introduction of a piece rate incentive scheme. Studying the performance of fruit-pickers, Bandiera, Barankay and Rasul (2005) report a 50% productivity increase as a response to a shift from relative performance incentives to piece-rate pay. They argue that under relative incentives good workers retain performance to favor their less able colleagues. For the same data, Bandiera, Barankay and Rasul (2009) find evidence that managers prefer workers they are socially connected with as long as they are paid fixed wages. However, when receiving incentives based on their subordinates performance, managers seem to favor the most able workers regardless of any social ties. Kahn, Silva and Ziliak (2001) study the introduction of a performance incentive program that compensates Brazilian tax officials for collecting taxes and uncovering tax violation. They find a 75% increase in fines per incident once the pay for performance program had been launched. Likewise, Fernie and Metcalf (1999) confirm the incentivizing effects of performance pay for jockeys. Based on an experiment on schoolteachers, Lavy (2009) reports teachers to respond to pay for performance schemes since pupils' academic achievements (e.g. conditional pass rate, mean test scores) improved significantly when teachers received bonus payments based on their classes' performance. Further evidence on the price effect following the introduction of exogenous incentives is given among others by Booth and Frank (1999) as well as Cadsby, Fei and Tapon (2007). The same pattern holds true for changes in incentive schemes that are unfavorable to employees: Freeman and Kleiner (2005) observe a significant decrease in productivity on the firm-level after one of the last shoe manufacturers in the US switched from piece rate pay to fixed pay.

[38] The other half is attributable to self-selection effects since the company is able to attract a more qualified workforce due the new incentive scheme.

In addition to studies on individual incentives, there is large support for the existence of a price effect at the group- or company-level, too.[39] For instance, fairly recent evidence is provided by Lavy (2002) on schools and student performance, Knez and Simester (2001) on company-wide incentives at Continental Airline, Jones, Kalmi and Kauhanen (2010) on production lines of a Finnish food-processing firm, Jones and Kato (1995) on bonuses and stock option plans at 109 Japanese companies, Boning, Ichniowski and Shaw (2007) on production lines in US minimills and more general Zenger and Marshall (2000) on the determinants of incentive intensity of team incentives as well as Che and Yoo (2001) on the general power of group-incentives.

Despite the broad support for the productivity enhancing effects of incentives and performance pay at the individual-level and the group-level (see Gibbons (1997) and Prendergast (1999) for detailed overviews), there are some difficulties in terms of hidden economic costs associated with performance incentives, too. First, individual incentives at least partially shift the risk from employers on to employees. In order to still meet the employees' participation constraints, employers need to compensate employees for taking the risk (e.g. Oyer, Schaefer 2011). As a consequence, overall labor costs increase to the disadvantage of firm profitability. Second, monetary incentives might shift employees' attention from task activity towards task outcome. Moreover, employees might focus only on those tasks whose outcomes are actually rewarded and put less or none effort on other tasks (e.g. Holmström, Milgrom 1991, Baker 1992). Overall performance might suffer. Third, based on data on Canadian tree planters, Paarsch and Shearer (2000) find that at least some part of an incentive-induced quantity increase comes at the expense of a reduction in quality. This argument is confirmed among others by Frick, Götzen and Simmons (2013) who observe German steel workers to increase output after the introduction of monetary incentives by means of reducing product quality. Fourth, further hidden costs may arise from cheating and shirking employees that aim to fully benefit from the incentive. For instance, Jacob and Levitt (2003) show that at least some schoolteachers cheat on their students' test scores when incentivized for good class results. To avoid shirking, it seems to be important to examine variables that may interact with financial rewards in affecting task performance. This argument suggests incentive pay plans to work best when matched with com-

[39] Although group-based incentives are prone to free-riding by employees (Holmström 1982), peer pressure and mutual monitoring might offset free-riding at least in small groups (e.g. Kandel, Lazear 1992). However, we do not intend to discuss group-based incentives in this study in detail.

plementary HRM practices (e.g. Jones, Kalmi, Kauhanen 2010, Ichniowski, Shaw, Prennushi 1997).

However, in recent years starting with the now seminal work by Deci (1971) psychologists have raised doubts on the benefits of incentives as praised by agency theory. Based on lab experiments on puzzle tasks as well as a field experiment on a campus newspaper's headline staff, Deci (1971) finds intrinsic motivation to decrease as soon as monetary incentives come into play. He concludes that financial rewards buy off – or crowd out – intrinsic motivation. Within the last decades, more and more economists have addressed this issue and presented extensive experimental and empirical evidence on the existence of the crowding-out effect. Hence, a new strand of theory has evolved that Gibbons and Roberts (2013, p. 90) refer to as behavioral incentive theory. The behavioral approach to incentive theory assumes peoples' actions to be not entirely attributable to material motives. Rather, they are a multidimensional consequence of intrinsic, extrinsic and reputational incentives (e.g. Bénabou, Tirole 2006). However, the introduction of extrinsic incentives may prove detrimental to overall employee effort and performance. Since extrinsic incentives may crowd out other motives that are also important in determining employee effort, extrinsic incentives may backfire and undermine the very behavior they intend to encourage.[40] As a consequence, with extrinsic incentives in practice employee productivity may be lower in comparison to a non-incentive environment. Put differently, monetary incentives that are often framed by economists as positive reinforcers of employee behavior may rather function as negative reinforcers (Bénabou, Tirole 2003).

In the literature, there are two main reasons discussed that may trigger the crowding-out effect of exogenous incentives (e.g. Gneezy, Meier, Rey-Biel 2011). First, the principal can be assumed to be best informed about the peculiarities of a contracted task. When setting incentives he should value both the task characteristics as well as the presumed ability of the agent to fulfill the task. Given these considerations, the introduction of an incentive might signal negative task characteristics to the agent and reduce his intrinsic motivation (e.g. Bénabou, Tirole 2003, Gneezy, Meier, Rey-Biel 2011). Furthermore, agents might sense a lack of trust if the principal feels the need to incentivize them extrinsically. Feeling or even being controlled is likewise found to reduce intrinsic motivation (e.g. Deci, Ryan 1985, Falk, Kosfeld 2006). Second, following the multidimensional construct of employee

[40] A substantial body of literature extends the theoretical considerations on this issue (e.g. Besley, Ghatak 2005, Ellingsen, Johannesson 2008, Prendergast 2008).

motivation proposed by Bénabou and Tirole (2006), one can expect extrinsic incentives to undermine other intrinsic or social incentives that motivate employees. A related body of work discusses crowding-out to arise among others from social preferences such as altruism (e.g. Frey 1994), social concerns about one's image, reputation and status (e.g. Ariely, Bracha, Meier 2009, Bénabou, Tirole 2006, Ellingsen, Johannesson 2008), the desire to reciprocate (e.g. Fehr, Falk 2002), trust (e.g. Sliwka 2007), effort norms and peer effects (e.g. Falk, Ichino 2006, Mas, Moretti 2009), but also individual values, beliefs and goals (e.g. Young, Beckman, Baker 2012), and the task itself (e.g. Frey 1997).[41] Hence, for both reasons, monetary incentives are expected to negatively affect an agents' intrinsic or social motivation and, as a consequence, reduce worker productivity.

The theory is backed by broad empirical evidence that confirms the existence of the crowding-out effect. One of the first related contributions is work by Titmuss (1970) who argues that paying monetary incentives to blood donors reduces blood supply since exogenous rewards crowd out other motives. This argument was confirmed in later studies by Mellström and Johannesson (2008) as well as Lacetera and Macis (2010). Based on two experiments, Gneezy and Rustichini (2000a) state that offering monetary incentives to individuals decreases their individual performance. First, students that are offered monetary incentives for each correct answer in IQ-questionnaires are performing worse unless the money paid per correct answer reaches a high enough level. This implies that performance incentives need to exceed some critical amount to do any good.[42] Second, students that are rewarded financially for voluntary work reduce their performance. According to the authors the material incentives replace students' intrinsic motivation as well as their motivation through social approval. Frey, Oberholzer-Gee and Eichenberger (1996) as well as Frey and Oberholzer-Gee (1997) survey Swiss citizens confronted with governmental plans of regional nuclear waste repositories. They observe citizens' support for the repositories to drop by half when offered public compensation. This emphasizes the argument that agents interpret incentives as compensation for negative consequences only the principal is aware of. Furthermore, Ariely, Bracha and Meier (2009) find monetary incentives to crowd out image concerns. They argue that people do not respond to highly visible extrin-

[41] Further empirical evidence on non-monetary incentives is extensive (e.g. Charness, 2004, DellaVigna, List, Malmendier 2012, Falk 2007).

[42] This result finds support in a recent study on contingent and non-contingent incentives. Surveys send to respondents either contained a reward or only promised a reward. Again, (contingent) incentives that are too small may undermine the intended behavior (Gneezy, Rey-Biel 2014).

sic incentives for prosocial behavior since this might be interpreted as pursuing self-interests by their peers.

Moreover, a substantial body of the literature suggests that negative impacts of monetary incentives are not limited to measurable performance. In experiments settled around a gift-exchange game, Fehr and Gächter (2002) provide evidence that voluntary cooperation is largely undermined in incentive contract situations in comparison to non-incentive contract situations.[43] Burks, Carpenter, Goette (2009) confirm this argument for Swiss bicycle messengers. Working with incentives in terms of penalties, Fehr and List (2004) conduct trust game experiments on both CEOs and students. They find that incentives penalizing shirking behavior may entail hidden costs in terms of less trustworthy behavior. In a field study, Gneezy and Rustichini (2000b) analyze the introduction of a monetary fine designed to incentivize parents to pick up their children from a private day-care center on time. In contrast to the expectations, late pickups significantly increased to twice the initial size after the fine was introduced. Again, incentives did not yield the desired effect. Further evidence on crowding-out of extrinsic incentives either in form of rewards or fines is given among others by Carpenter and Myers (2010), Georgellis, Iossa and Tabvuma (2011) as well as Reeson and Tisdell (2008). Several meta-analyses further support the crowding-out effect of monetary incentives (e.g. Deci, Koestner, Ryan 1999, Rummel, Feinberg 1988, Tang, Hall 1995, Wiersma 1992).

In summary, classic personnel economics theory expects monetary incentives to align interests of principals and agents and, therefore, to increase employee effort choice decisions and performance. In contrast to this assumption, behavioral incentive theory suggests extrinsic rewards to crowd out other intrinsic and social incentives that are equally important in motivating employees. As a consequence, overall employee effort and performance might be lower under extrinsic incentives. Given both theoretical assumptions, our aim is to analyze how incentives may affect worker performance and productivity. In the organizational setting at hand, commercial truck drivers were incentivized by the introduction of a monetary reward for meeting the company's performance targets. After two years in practice, the incentive scheme was abolished by the company. If classic economic theory holds true we would expect a better performance while incentives were in practice due to the price effect of exogenous rewards. Thus, we should observe a decrease in worker per-

[43] This finding is not limited to financial rewards. Drago and Garvey (1998) tested helping behavior at work and observed a negative link between promotion incentives and helping behavior.

formance and productivity after the monetary incentives are abolished. However, if behavioral incentive theory proves true, we expect to find a reduced performance while bonus pay is in practice due to crowding-out. In other words, we should observe performance to increase after the incentives are abolished since drivers might draw their motivation from other intrinsic or social motives when no longer financially incentivized.

4.3 Data Set and Descriptive Statistics

In this paper we seek to analyze behavioral responses of workers to shifts from a bonus pay scheme to a non-incentive environment. Therefore, we use a hitherto unavailable panel data set that includes trip-wise performance information of commercial long-haul truck drivers of an in-house hauler of a large European truck manufacturer. The data is based on information recorded by GPS-based in-vehicle computers of the hauler's fleet management system. Performance data covers an extended period of thirty-six months (January 2011 to December 2013). With the beginning of the observation window, the hauler installed a monetary performance premium to reward good driver performance. However, after two years in practice, the hauler decided to disestablish the incentive scheme again.

Within the organizational structure of the parent company, the hauler realizes just-in-time delivery of intermediate goods from one production plant of the truck manufacturer to another. While on the road, drivers travel roughly 1,250 kilometers across four Northern European countries. To comply with drivers' mandatory resting periods[44], the total distance is split up into two stages with a fixed location for the drivers to rest. At this location, the hauler operates a premise that provides sleeping, cooking and sporting facilities for drivers during their rest. Upon their arrival at the location, drivers pass on their trucks to another driver who continues the journey on the very same truck. In other words, driver A arrives at the resting location, where driver B has just finished his mandatory rest. Driver B takes over the truck and immediately drives on to the final destination, whereas driver A stays for his rest and awaits the arrival of the next truck to continue his journey. Thus, drivers travel a round-trip with four stages on four different trucks before returning home. This procedure secures just-in-time supply of goods without transferring the company's ware-

[44] Due to European road haulage regulations drivers are not allowed to exceed 9 hours of driving per day with a subsequent mandatory rest period of at least 11 hours. There are few exceptions enabling drivers to break these regulations in order to react to current traffic conditions or short-term workload peaks (European Commission 2014b).

house on to the streets. Obviously, driver-truck sorting in this setting is totally random since drivers are required to take the next truck available without having any choice.

Our initial data set includes 81 commercial truck drivers resulting in n=11,439 driver-trip observations.[45] We focus exclusively on drivers on permanent contracts to avoid confounding effects attributable to incentives arising from temporary contract characteristics as observed in other studies (e.g. chapter five of this thesis, Bradley, Green, Leeves 2012). All trucks are reported to be deployed almost exclusively on long-haul assignments from one production facility to the other. To fully exclude short-haul assignments we restrict the data to trips with a length between 520 kilometers and 720 kilometers. We selected this range since the two production facilities are 570 km respectively 670 km away from the resting facility.[46] Moreover, we exclude those observations with a reported fuel consumption outside the range of minus two times and plus four times the standard deviation around the mean. Finally, we only include those drivers that we can observe at least on 80 trips during the observation period (i.e. approximately four to five months of regular driving) and that have experienced the incentive scheme for at least three months. These constraints reduce the initial dataset thirty-seven truck drivers and n=6,825 observations.[47]

We merged the trip-wise performance information with demographic data of drivers (age, gender and tenure). It appears from the summary statistics presented in Table 4.1 that the sample is dominated by male drivers, a fact that comes as no surprise in the trucking industry. Still, three female drivers are included in the data accounting for 580 observations (8.5% of total observations). Reported driver age covers almost the entire working life span from 22 to 64 years (mean=44.08, s.d.=11.07). This variance allows for a thorough analysis of potential age effects. Tenure as a proxy for driving experience is included on a monthly level given the exact date a driver joined the company (mean=25.59 months, s.d.=12.02). Since we exclude drivers from the sample that are employed on a temporary contract, employment status is similar for all drivers. Moreover, we could neglect any cul-

45 The panel data set is unbalanced due to driver and vehicle turnover during the observation period. For instance, at this company trucks are replaced after a total mileage of 1m kilometers.

46 We extended both distances by +/- 50km to account for drivers who exceed the maximum allowed driving time before reaching their destination or need to take alternative routes due to construction work or traffic congestion.

47 However, not all models could be estimated with the full data set since not all vehicles are equipped with the latest version of the relevant on-board communication units. Hence, incomplete information produces missing values at some instances.

tural differences due to the fact that all drivers originate from the country domestic to the hauler.

Table 4.1: Summary Statistics - Driver Demographics

Variable	mean	s.d.	min.	max.
Driver Age (in years)	44.08	11.07	22	64
Driver Gender (male=1)	.915	-	0	1
Driver Tenure (in month)	25.59	12.02	1	64
Bonus (Bonus=1)	.58	-	0	1
Number of drivers: 37 Number of driver-trip observations: 6,825				

The GPS-based in-vehicle computers provide the hauler with extensive real-time information offering considerable opportunities to manage and monitor fleet activities including truck positioning, fuel consumption, driving behavior, truck handling and vehicle status. Table 4.2 presents the fuel consumption in liters per 100 km (mean=27.89 l, s.d.=3.02 l) as well as a shortlist of the most important driving variables (e.g. brake applications, speeding, average speed, maximum speed).

Table 4.2: Summary Statistics - Driving Style

Variable	mean	s.d.	min.	max.
Fuel Consumption (in l/100 km)	27.89	3.02	21.2	39
Average Speed (in km/h)	75.32	3.34	42.1	84.7
Idling (in % of engine running time)	.023	.02	.002	.238
Coasting (in % of engine running time)	.14	.05	.007	.399
Speeding (in % of engine running time)	.079	.11	0	.892
Brake Applications (# per 100km)	12.62	7.83	.2	149.1
Harsh Brake Applications (# per 100km)	.11	.25	0	10.1
Harsh Accelerations (# per 100km)	.07	.27	0	4.3
Maximum Vehicle Speed (in km/h)	96.23	5.55	84	114
Number of drivers: 37 Number of driver-trip observations: 6,825				

The descriptive data already reveal interesting insights. First, idling – which is often labeled to be one of the most important determinants of fuel combustion (e.g. Lutsey et al. 2004) – is not of major relevance in the organizational setting at hand (mean=2.3% of engine running time). This might be due to the fact that drivers spend most of their time on highways and rarely face stop-and-go traffic in cities. Second, drivers are required by national legislation of the transit countries to drive 80 km/h at max. However, maximum values for average speed (max=84.7 km/h) and maximum vehicle speed (max=114 km/h) suggest that drivers override national speed limits occasionally. This might indicate that at times drivers need to catch up for traffic-induced delays to secure just-in-time supply. Third, the simple min-max comparison of several driving parameters reveals a high heterogeneity (e.g. brake applications vary from .2 to 149.1 per 100 km). This indicates that driving style might to a large degree be influenced by exogenous effects such as traffic density or weather conditions. To address these issues we add external information to the data that enables us to control for exogenous conditions. All controls will be discussed in detail in the next section.

Based on the information recorded by its in-vehicle devices, the fleet management software is programmed to compute scores that assess driver performance in relation to target values predefined by the hauler. These scores cover drivers' anticipation behavior, choice of gear, use of brakes and hill driving behavior.[48] In addition, the software uses these scores as well as further driving parameters to assess overall driver performance for each trip as being either "good", "mediocre" or "poor". Expressed by a green-yellow-red traffic light visualization, the trip evaluation is reported back to the hauler via the fleet management system. For transparency reasons, all results are accessible for all drivers via an online tool. Although fuel consumption is not directly evaluated by the software, a "good" driving behavior is congruent with an eco-friendly driving style. It appears from Table 4.3 that driving scores on average display high means varying from 63% to 81%. Yet, we question the choice of gear score since a mean of 97% appears to arise from the usage of automatic transmission vehicles in the majority of the fleet. The limited number of observations on this score fits this picture. Concerning the traffic light evaluation scores, exactly half of all trips are evaluated as being "good" (50%), while slightly fewer trips are reported as being "poor" (45%). Only a small number is rated as being "mediocre" (5%).

48 The underlying algorithms that assess driver performance and generate scores are corporate secrets and were not revealed to us.

Table 4.3: Summary Statistics - Traffic Light Performance Evaluation Tool

Variable	mean	s.d.	min.	max.	Observations*
Driving Scores					
Anticipation (Score in %/100)	.81	.16	0	1	5,251
Choice Gear (Score in %/100)	.97	.14	0	1	3,673
Use Brakes (Score in %/100)	.77	.19	.1	1	5,064
Hill Drive (Score in %/100)	.63	.25	0	1	5,028
Trip Evaluation					
Overall (1=green, 2=yellow, 3=red)	1.94	.97	1	3	6,801
Good (yes=1)	.50	-	0	1	-
Mediocre (yes=1)	.05	-	0	1	-
Poor (yes=1)	.45	-	0	1	-
Number of drivers: 37					

*Varying number of observations since not all on-board computers report complete data at all instances.

For an overview of mean, standard deviation, as well as within and between variance of all dependent variables listed separately for the period with incentives in practice and the period after incentives have been abolished, see Table A.6 in the appendix.

Our study design has numerous advantages and contributes to personnel economics in several regards. First, the organizational setting allows studying the effects of the abolition of incentives. As mentioned before, the hauler installed a performance bonus as of January 1st 2011. Based on the evaluation scores reported by the fleet management software, drivers were rewarded financially for good performances. We like to point out again that low fuel consumption is only indirectly rewarded. The maximum amount a driver could earn on incentives did not exceed 5% of his monthly fix pay. Two years later, the bonus system was abolished by the hauler at December 31st 2012. Interestingly, the hauler installed the performance bonus without telling the drivers for the first three months. Despite the delayed disclosure, we doubt that the introduction of the incentive scheme remained completely unnoticed by drivers without any rumors spreading or premature information leaks. This would bias any findings. Thus, we focus our main estimations exclusively on the abolition of the bonus at the end of December 2012. This seems justified since there is broad evidence on productivity effects following the installation of incentives (e.g. Lazear 2000b, Shearer 2004), but to our knowledge we are among the first researchers who work on the abolition of an incentive scheme in practice, with Freeman and Kleiner (2005) being a rare exception. Our set up, therefore, might offer new insights on the functioning of incentives that have not been covered by academic research so far. Second, comparisons among truck drivers are often hampered due to inconsistent exogenous influences coming along with different assignments, destinations and routes. Since the round-trip in the organizational setting at hand is identical for all drivers and constant over time, we can assume each driver to be exposed to comparable exogenous influences while on the road (e.g. road condition, constructions, detours etc.). This allows for comprehensive analyses of driver performance. Third, driver-truck sorting is totally random in the organizational setting of this study since drivers cannot influence which truck is the next available at either stage of their round-trip. Thus, we are able to control for good or poor driver-truck matches. Fourth, we can neglect any selection bias of agents sorting into specific incentive schemes since most drivers were already employed prior to the installation of the incentive scheme. Fifth, detailed date and time information for each trip allows us to control for exogenous effects (e.g. weather conditions) and temporary traffic density due to holiday traffic or daily rush hours. Both weather and traffic density can be assumed to highly determine truck handling and driver performance. Finally, vehicle load factors need to be similar over time to avoid biases arising from unequal vehicle maneuverability. Data by the European Commission (2014c) suggest that almost one quarter of all vehicle-kilometers of trucks within the EU is run empty.[49] However, the hauler reports average load factors above 90%, thus we can assume similar cargo weights and equal maneuverability of trucks for all trips. This allows us to neglect this issue for our data.

4.4 Empirical Models and Estimation Results

Our data offers broad opportunities to measure truck driver performance. In total, our estimations are based on four dependent variables that proxy for drivers' effort choice and productivity. First, we use the average fuel consumption in liters per 100 kilometers. Although low fuel consumption is not directly incentivized, all target values set with regard to driving parameters and performance evaluations are based on eco-friendly driving behavior and result in low fuel use. Since drivers receive individual fuel efficiency training on a non-regular basis, we can assume them to be well informed on how to meet company

[49] According to the European Commission (2014c) this is mainly attributable to strict cross-border transit regulations.

standards.[50] Drivers deviating from these standards, therefore, may be assumed to have chosen low effort levels. Thus, we are convinced that fuel consumption used as dependent variable may serve as a good proxy for drivers' effort choice decisions. In our sample, average fuel consumption is 27.89 l/100 km with a standard deviation of 3.02 liters (see Table 4.2). Second, we deploy the most important driving parameters such as speeding, brake applications and average speed (see Table 4.2). Due to fuel efficiency training, drivers know how to meet company standards. Differing values, therefore, may again be an appropriate proxy for driver effort choice. Third, we use the four main driving scores (anticipation, use of brakes, choice of gear, hill drive) recorded by the in-vehicle computers (Table 4.3). These scores directly influence trip evaluation and are highly relevant in determining the amount of bonus paid. Full online excess to the fleet management data base should raise drivers' awareness on how they performed on each score. Hence, we assume drivers to know their individual strengths and weaknesses in driving behavior quite accurately. Otherwise put, drivers are aware of how they can increase individual bonus pay by improving on those scores that do not yet meet company standards. In our estimations, we model low driving scores to proxy for low effort choices of drivers. Fourth, we deploy internal traffic light evaluation results based on software calculations (see Table 4.3). Again, drivers are offered full online excess to the evaluation reports that rate trips as either "good", "mediocre" or "poor". In other words, drivers know about their individual performance and should be aware of any misconduct.[51] Since they have both the ability and the knowledge to drive according to company standards, we believe trip evaluations other than "good" to result from low effort choice decisions. All four performance measures are insensitive to any subjective rater bias since they are recorded and assessed by in-vehicle computers automatically. While fuel consumption, driving parameters and evaluation scores are categorical variables, trip evaluation is of ordinal nature.

The estimated models have the following general form (with *FUELC* being the fuel consumption, *DRIVPAR* representing the driving parameters, *SCORE* representing the evaluation scores and *EVAL* being the traffic light evaluation reports):

---

[50] According to van Mierlo et al (2004) (car) drivers know by intuition how to drive eco-friendly. Thus, it does not really matter that we have no information neither on the frequency nor the exact dates of individual fuel efficiency training since we believe any training effect to arise from recollection instead of new insights.

[51] According to Huang et al. (2005) and Roetting et al. (2003), truck drivers appreciate feedback from technology.

$$\left.\begin{array}{l} FUELC \\ DRIVPAR \\ SCORE \\ EVAL \end{array}\right\} = \alpha + \beta DRIV + \gamma VEH + \delta TRAF + \varphi WEA + \omega CONTR + \varepsilon \quad \begin{array}{r} (1) \\ (2) \\ (3) \\ (4) \end{array}$$

where α is the constant, *DRIV* describes driver performance as expressed by equation (5), *VEH* controls for vehicle effects, *TRAF* describes a vector of external traffic conditions, *WEA* describes a vector of weather conditions, *CONTR* is a vector combining further controls, β, γ, δ, φ and ω are the estimated coefficients and ε is the error term. The independent variables are discussed in more detail below.

Driver performance (*DRIV)* is modeled as a function of the following form:

$$DRIV = f\,((AGE + HUMC) * EFCH) \qquad (5)$$

$$\text{with} \quad EFCH = Intrinsic\ Incentives + Extrinsic\ Incentives \qquad (6)$$

where *AGE* measures driver age and *HUMC* proxies driver human capital as driving experience with the hauler (monthly tenure). Age and experience are important personal characteristics in determining driver performance. On the one hand, age has been proven to be detrimental to important abilities relevant to driving. For example, older drivers display an extended response time as well as lower visual and psychomotor abilities (e.g. Llaneras et al. 1998). On the other hand, experience is found to compensate these age-related impairments (e.g. Brock, Llaneras, Swezey 1996, Guest, Boggess, Duke 2014). The importance of experience is even more pronounced when taking accident risk into account since younger and less experienced drivers have a higher risk to be involved in accidents (e.g. Häkkänen, Summala 2001, McCall, Horwitz 2005, Rodriguez, Targa, Belzer 2006). Thus, any analysis without including age and experience into the estimations would produce biased results. In lack of information on skills and qualification, we interpret age and experience as the overall performance potential of employees. However, the performance potential is at any moment subject to a drivers' individual effort choice decision (*EFCH*). Performance potential and effort choice eventually translate into actual driver performance. We model *EFCH* to be the joint effect of both intrinsic and extrinsic performance incentive (see equation (6)). While in the organizational setting at hand intrinsic incentives might include motives such as social preferences, altruism or eco-friendliness, extrinsic incentives comprise the monetary performance pay incentives.

We use further variables to control for exogenous effects that might influence truck handling and driving behavior as modeled in equations (1) to (4). First, we use a vehicle dummy variable *VEH* to control for different truck configurations. A different set of technological equipment might facilitate truck handling at some instances or influence fuel consumption solely for technological reasons. Second, we use a vector of date and time variables *TRAF* that proxy for traffic density while on the road. Concerning the time of the day, we assume rush hours in the morning (7 to 9 a.m.) and evening (4 to 6 p.m.) to impose high traffic density, while drivers being on the road at night might benefit from open highways. Dummy variables control for rush hours and trips at night alike (trips from 8 p.m. to 7 a.m.). Additionally, we control for day of the week effects since we expect e.g. Mondays and Fridays to be prone to congested roads due to weekend commuters. Furthermore, we include a dummy variable covering a three-day window around international holidays since we believe roads to be congested due to holiday traffic at these dates.[52] Third, a vector of variables, *WEA*, controls for weather conditions during each trip. This seem necessary since previous research found weather to highly affect driving behavior, general traffic conditions and crash risks (e.g. Kilpeläinen, Summala 2007, Norrman, Eriksson, Lindqvist 2000). We, therefore, add control variables for temperature, precipitation, snow depth and wind speed.[53] Eventually, we apply dummy variables for all thirty-six months of the observation period to account for possible trend or timing effects. Likewise, we use dummy variables for all fifty-two weeks of a given year to account for seasonality. Taken together, this broad range of control variables excludes the major sources of potential exogenous influences on driver performance. Thus, we are confident that any performance effect that we will detect in our estimations is entirely attributable to drivers' effort choice decisions as a response to the incentive scheme. In the following, we discuss results on fuel consumption as modeled by equation (1) first, then report our findings on driving parameters as modeled by equation (2). Afterwards, we present results on driving scores as modeled by equation (3) and eventually discuss the estimations on trip evaluation as modeled by equation (4). The key estimations are reported in Tables 4.4 to 4.7, respectively. Overall, our results support the crowding-out effect of extrinsic incentives since truck

[52] Only those holidays that are mandatory in all four countries drivers pass through during their round trips are considered (e.g. New Year's Day, Easter, Christmas, etc.).

[53] Since we do not have detailed truck positioning data, we gathered weather information at a meteorological station approximately halfway the total distance. This should level differing climate conditions at the starting and arrival points. Data was provided by the German Meteorological Service (DWD 2014), a German state institution.

drivers' effort (and as a consequence their performance and productivity) is significantly lower with monetary performance incentives in practice.

First visual evidence on divergent levels of fuel consumption with and without incentives in practice is presented in Figure A.4.1 in the appendix. Comparing the kernel density functions for both incentive environments, it appears that fuel consumption shifts to the left after incentives have been abolished. This indicates lower fuel use without incentives in practice. As a first empirical test for significant behavioral responses, we run a two-way ANOVA on the thirty-seven drivers from April 2011 to December 2013 to examine the effect of incentives on fuel consumption as a proxy for driver performance and productivity (see Table A. 7 in the appendix). We focus on a time frame starting in April 2011 since the first three months of the observation window might be biased due to the secret introduction of the performance pay.[54] We found a significant effect of the bonus pay on driver's fuel consumption indicating that significantly different quantities of fuel are used depending on the existence and non-existence of monetary incentives, $F(1, 6{,}252)=4.31$, $p=.0379$. Since the ANOVA does not report any significant differences in the simple comparison of drivers ($p=0.5263$) or the driver-bonus interaction ($p=0.4610$), we are confident that any performance effect estimated in this study may be explained by behavioral responses entirely attributable to the (non-)existence of the incentive scheme. Moreover, this finding supports our argument that drivers adapt their individual effort choice to the incentive environment. To follow these indications, we apply elaborate empirical estimations to our data. We start with fixed-effect specifications since we are particularly interested in within-driver behavioral responses to the disestablishment of the incentive scheme. Moreover, fixed-effects models account for present driver fixed-effects that are not reported in the data. These effects can be either time-invariant (e.g. talent, prior driving experience, etc.) or quasi time-invariant without any or only slight modifications over time (e.g. driving style, environmental friendliness, professional ethos, etc.). To account for these unobserved effects that are seldom uncorrelated with explanatory variables, fixed-effect specifications are preferable (e.g. Wooldridge 2013). Since the incentive scheme was introduced without telling the drivers for the first three months we restrict our estimations to a period from April 2011 to December 2013 to avoid any bias arising from premature information leaks.

[54] The secret installation might not have been as secret as intended by the hauler since we assume rumors to have circulated among the drivers prior to the introduction of the new remuneration scheme.

Table 4.4 displays the baseline results of the fixed-effect estimations with fuel consumption as the dependent variable (variable *FUELC* in equation (1)). A total number of thirty-seven drivers and a mean of 171 observations per driver (min=80, max=221) allow for profound fixed-effect estimations. We step-wise built up our estimations using varying model specifications to detect potential biases that may arise from including control variables. Our preferred model specification is the full model (6). A modified Wald test for the full model specification allows rejecting group-wise heteroskedasticity (Prob>chi2=.8305).

Table 4.4: Baseline Results of the Fixed-Effects Estimations (Models 1 to 6)[+]

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Incentives (yes=1)	.166**	.324***	.329***	.392*	.454**	.419**
	(.0764)	(.110)	(.113)	(.199)	(.180)	(.191)
Age (in years)				.0506		-.447
				(.114)		(.287)
Tenure (in months)					.00842	.0436*
					(.00859)	(.0245)
Constant	27.77***	26.89***	26.87***	24.59***	26.60***	45.62***
	(.0476)	(.321)	(.414)	(5.233)	(.528)	(12.22)
Vehicle Controls	YES	YES	YES	YES	YES	YES
Time Controls		YES	YES	YES	YES	YES
Date Controls			YES	YES	YES	YES
Weather Controls			YES	YES	YES	YES
Driver FE	YES	YES	YES	YES	YES	YES
Observations	6,326	6,326	6,326	6,326	6,326	6,326
R-squared	.001	.012	.017	.017	.017	.018
Number of driver	37	37	37	37	37	37

Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.1
[+]April 2011 to December 2013; dependent variable: fuel consumption.

The fixed-effect estimations seem to support the arguments of the crowding-out theory. We find behavioral responses to the disestablishment of the incentive scheme to be statistically significant at the 5%-level: Drivers chose less effort and, therefore, use more fuel when monetarily incentivized. Thus, incentives that should motivate drivers to eco-drive are in fact backfiring on the very behavior they intend to stimulate. Other important intrinsic motives that might encourage drivers to eco-drive (e.g. environmental believes, professional ethos, peer group competition) seem to be suppressed by financial rewards. As a consequence, overall performance decreases and net fuel consumption increases. As soon as the extrinsic incentives are abolished, these motives seem to revive and enhance driver performance. The estimated coefficient of the incentive effect compared to the non-incentive situation is .419 (standard error=.419) in our preferred full model specification (model 6). Thus, ceteris paribus, a driver uses .419 l/100 km more fuel with incentives in practice. In other words, in reducing driver performance and productivity incentives pose an extra burden to the hauler's baseline results. Since the average driver is on the road for about 125,750 kilometers per year and consumes approximately 35,000 liters of fuel while on the road, the additional fuel use attributable to the extrinsic incentive adds up to roughly 530 liters per driver and year. Given an average fuel price of €1.50 per liter, this amounts to hidden costs of incentives of around €800 per driver and year that have to be added to the net costs of the incentive itself. Thus, the incentives appear disadvantageous for the hauler in two ways.

In contrast to previous research (e.g. Rodriguez, Targa, Belzer 2006), human capital in terms of driver experience – proxied by cumulated tenure with the current employer – is not improving performance. Instead, each month a driver stays with the hauler seems to be even more detrimental to performance with fuel use increasing by .04 liters per 100 km. Yet, our findings do not allow for long-perspective reasoning since maximum tenure of drivers in our data set is limited to 64 months. Thus, we believe greater variance in tenure would be necessary to meaningfully interpret this finding. Concerning age, we are not able to detect an effect in any of our estimations. Referring to equation (1) and (2), both the negative tenure effect and the missing age effect emphasize the importance of motivational effort choices (*EFCH*) in determining performance at least in a low-skill occupation such as trucking. Interestingly, we find significant vehicle effects for some trucks. This is especially surprising since the hauler replaces trucks regularly after reaching a total mileage of 1 million kilometers which usually happens every two years. Thus, the hauler's fleet should only consist of similar state-of-the-art trucks. One possible explanation may be found in the particular role of the hauler within the truck manufacturer's organization. Certain trucks of the fleet are occasionally equipped with newest technology developments for real-life trial e.g. endurance tests. Thus, although trucks seem to be equal at first glance they may be not. However, we do not have any information on which trucks are equipped and which effect on fuel consumption or driving behavior can be expected by the modifications. Yet, this should not be an issue in our estimations since drivers are observed on four different trucks per round-trip resulting in various driver-truck matches. Fairly unexpected, neither day of the week nor time of the day show any statistically significant influence on fuel consumption.[55] Significantly more fuel is used only during the three-day window around international holidays (+.42 l/100 km, significant at the 5%-level). This indicates that around holidays driving indeed is impeded by traffic density. Eventually, concerning weather conditions our results oppose expectations raised by previous research on the impact of weather on driving (e.g. Kilpeläinen, Summala 2007, Norrman, Eriksson, Lindqvist 2000). We only find air temperature to influence fuel use as overall fuel consumption is reduced by -.04 l/100 km with each additional degree Celsius. This might be related to the often freezing temperatures in Northern Europe including glazed frost and snow falls which turn truck handling more difficult and require more fuel due to frictional resistance.

To check for the robustness of our findings and to better understand the timing of the incentive effect, we re-estimate the fixed effect models on fuel consumption using different periods of time. First, we address doubts on the secret introduction of the performance incentives by estimating fixed-effect models for the first eight months of the observation window (January 2011 to August 2011). That is, three months during which drivers were not aware of the introduction of incentives (treated as 0 in the estimations) and the five consecutive months when drivers were informed about the existence of the bonus scheme (treated as 1). In support of our baseline results, estimations on the full model specification (see model (5) of Table A.8 in the appendix) show a strong initial effect of 2.223 liters used additionally per 100 km with incentives in practice (significant at the 10%-level). This is roughly five times the effect of our baseline findings for the entire period of April 2011 to December 2013. Second, we narrowed the observation window to a period from May 2012 to August 2013. Since this period covers eight months pre and post the abolition of performance bonuses, it allows for a closer examination of the effect of abolished incentives (see Table A.9 in the appendix for results). We prefer model (5) excluding age over the full model since age and bonus are highly correlated in this setting (both change only once on January 1st 2013). It appears from the estimations that fuel consumption is higher with incentives in practice. Again, this finding supports our baseline results with the coefficient being twice as high as our baseline results suggest (additional .964 liters per 100 km, significant at the 1%-level). In general, our findings so far support arguments of

55 However, the controls for rush hours in the mornings (+.06 l/100km) and evenings (+.11 l/100km) indicate the expected positive relationship yet without reaching statistical significance. In contrast, the control for trips at night shows a counterintuitive positive relationship (+.2 l/100km), again without reaching significance.

monetary incentives crowding out other intrinsic motives that are likewise important for driver motivation. Yet, it appears as if behavioral responses are more pronounced in the first months after a shift in the incentive scheme.

In a second step, we estimate seemingly unrelated regressions (SUR) with the most important driving parameters as dependent variables (vector *DRIVP* in equation (1)). SUR was first proposed by Zellner (1962) and represents a special case of the generalized regression model. It allows for the estimation of regression equations that each have its own dependent variable but share a common set of independent variables. Most likely, the error terms are correlated across equations in such models. This issue has to be accounted for when estimating these model specifications. Furthermore, estimating equations separately would neglect the important information that a similar set of regressors appears in all equations (Greene 2003). Since equation estimations for each driving parameter share a common set of independent variables, SUR is the most reasonable choice at this point. Results of SUR estimations are presented in Table 4.5. Again, we exclude the first three months of the observation period. It appears from the estimations that overall driving behavior seems to be more aggressive and less fuel efficient with the performance incentives in practice. More precisely, we find drivers to practice more speeding (plus 1.1 percentage points of total driving time), use the brake more often (plus 3.9 brake applications per 100 km), use the brake more often in a harsh manner (plus .03 harsh brake applications per 100 km) and make less use of coasting (minus 1.3 percentage points of total driving time). With the bonus pay abolished, the drivers display higher effort in terms of a more fuel efficient driving behavior. In other words, drivers do not seem to comply with eco-friendly company targets as long as they are financially rewarded. Instead, they respond to the abolition of the incentive scheme by choosing that effort level the hauler originally intended to encourage by paying incentives. Still, we do not find significant effects on maximal vehicle speed, average speed, idling and harsh accelerations. The explanation for this might be straightforward. While idling seems not to be an issue in highway carriage, both maximal and average vehicle speed levels are limited at some point due to technological (and legislative) reasons. Only harsh accelerations would have indicated aggressive driving behavior. The lack of significance at this point, thus, represents a counterintuitive finding.

Table 4.5: Seemingly Unrelated Regression (SUR) Results on Driving Parameters+

Variables	Speeding	Max Vehicle Speed	Harsh Brake Applications	Brake Applications	Harsh Accelerations	Coasting	Average Speed	Idling
Incentives (yes=1)	.0112* (.00652)	.493 (.313)	.0269* (.0148)	3.896*** (.449)	.00918 (.0153)	-.0138*** (.00281)	-.270 (.195)	-.00169 (.00108)
Age (in years)	.0105 (.0124)	-.292 (.595)	-.00215 (.0282)	1.209 (.852)	-.0382 (.0291)	-.00754 (.00534)	-.733** (.369)	.000277 (.00205)
Tenure (in months)	-.000443 (.00103)	.0953* (.0494)	.00445* (.00234)	.0793 (.0708)	.00455* (.00242)	-6.50e-05 (.000443)	.0740** (.0306)	-.000379** (.000170)
Constant	-.376 (.566)	107.9*** (27.21)	.122 (1.288)	-49.22 (38.98)	1.803 (1.331)	.508** (.244)	109.3*** (16.89)	.0202 (.0938)
Vehicle Controls	YES	YES	YES	YES	YES	YES	YES	YES
Date Controls	YES	YES	YES	YES	YES	YES	YES	YES
Time Controls	YES	YES	YES	YES	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES	YES	YES	YES	YES
Driver Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	6,326	6,326	6,326	6,326	6,326	6,326	6,326	6,326
R-squared	.051	.066	.044	.047	.017	.052	.027	.082

Standard errors in parentheses
*** p<.01, ** p<.05, * p<.1
+April 2011 to December 2013

Overall, our results provide great support for behavioral incentive theory and its crowding-out assumption: Monetary incentives do not yield the behavioral changes desired by the hauler.

Concerning the four major evaluation scores anticipation, choice of gear, use of brakes and hill drive we estimate a further SUR (vector *SCORE* in equation (1)). Again, we do so to account for correlated error terms across equations arising with similar sets of independent variables. Incomplete data restrict our estimations to a limited number of n=3,138 observations (April 2011 to December 2013). It appears from Table 4.6 that drivers score worse on the parameters (given in %/100) when subject to incentives.[56] In numbers, drivers score nine percentage points worse on hill driving, thirteen percentage points worse on anticipation behavior and sixteen percentage points worse on the use of brake score when being rewarded financially for good performance.

Table 4.6: Seemingly Unrelated Regression (SUR) Results on Evaluation Parameters+

Variables	Anticipation	Use of brakes	Hill drive
Incentives (yes=1)	-.133***	-.158***	-.0899***
	(.0108)	(.0142)	(.0200)
Age (in years)	-.0170	-.0317	-.000994
	(.0191)	(.0250)	(.0354)
Tenure (in months)	-.00407***	-.00515**	-.00411
	(.00158)	(.00207)	(.00293)
Constant	1.857**	2.561**	.816
	(.873)	(1.146)	(1.619)
Vehicle Controls	YES	YES	YES
Date Controls	YES	YES	YES
Time Controls	YES	YES	YES
Weather Controls	YES	YES	YES
Driver Controls	YES	YES	YES
Observations	3,138	3,138	3,138
R-squared	.203	.135	.094

Standard errors in parentheses
*** p<.01, ** p<.05, * p<.1
+April 2011 to December 2013

Similar to our previous findings this result supports the crowding-out effect of monetary incentives. After the abolition of the incentive scheme, driver effort in terms of performance scores improved significantly. As in most other estimations, age does not have any

[56] We deliberately leave out the choice of gear score from our considerations as a mean of .97 (see Table 3) suggests that the entire fleet consists of automatic transmission vehicles.

statistically significant influence on driver performance in any model specification. Yet, tenure exhibits a minor but significant impact on evaluation parameters with anticipation scores reducing by .4 percentage points and use of brakes score reducing by .5 percentage points with each month a driver is employed at the hauler. This result opposes theoretical considerations on human capital as an important determinant of employee performance (e.g. Rodriguez, Targa, Belzer 2006). We believe decreasing attentiveness as a result of routine to be one possible explanation since truck driving is a very monotonous occupation – especially in settings with constant routes and on highways. However, the limited observation window of only 33 months as well as the overall limited tenure (max of 64 months) might bias our findings at this point. Vehicle dummies only seldom reach statistical significance. Coefficients for weather condition as well as date and time dummies are inconsistent for all four variables. Therefore, we refrain from reporting these results in detail.

In a final step, we estimate models on the traffic light trip evaluation assessed by in-vehicle computers (variable *EVAL* in equation (1)). We use trip evaluation as a proxy for driver effort choice decisions (*EFCH*) and, thus, are confident to detect potential performance effects of incentive. Following Gujarati and Porter (2009) as well as Wooldridge (2013) we estimate ordered logit and ordered probit regression models to address the ordinal nature of trip evaluation data. However, since probit regression only constitutes a minor modification of logit regression and yields barely different findings (Greene 2003), we focus on logit regression results in this paper due to reasons of brevity. Table A.10 in the appendix displays results for the ordered logit estimations with trip evaluation being the dependent variable. It becomes obvious that evaluations are significantly worse when performance incentives are in practice. Post-estimation probabilities of trip evaluation with and without incentives in practices are given in Table 4.7.

Table 4.7: Estimated Probabilities for Trip Evaluation in both Incentive Environments

Variables	*Incentives in Practice*	*Incentives Abolished*	*Change*	*95% CI for Change*	
Trip Evaluation					
Good	.4034	.6894	.2860	-.3421	-.2298
Mediocre	.0619	.0513	-.0106	.0072	.0139
Poor	.5347	.2594	-.2754	.2219	.3289

With incentives in practice the probability of a trip being assessed as “good” by in-vehicle computers is 40.34% compared to 68.94% after the abolition of incentives. This is a difference of 28.6 percentage points or 70.89%, respectively. At the same time, the probability of a trip being evaluated as “poor” dropped by 27.54 percentage points or 51.5% from 53.47% to 25.94%. Probabilities for “mediocre” trip evaluations vary only slightly by 1.06 percentage points or 17.2% after the abolition of the incentives. In other words, with performance incentives in practice the probability of a good truck driver performance is markedly lower compared to a situation without any extrinsic incentives. This indicates drivers to choose lower levels of effort when monetarily incentivized. Similar to the other results of this paper, this finding supports the crowding-out assumption of extrinsic incentives backfiring on the very behavior they intend to prevent.

All estimations represented in Tables 4.5 to 4.7 are based on the full model specification. To control for biases arising from control variables we re-estimated all models under different specifications. Results proved to be robust for all estimations with little variation in coefficients. The same holds true when excluding the choice of gear score from the SUR models. Findings for the other dependent variables remain robust. We further check for the robustness of the results in two ways. First, we re-estimate our models excluding the three women from the data set. Second, we include all temporary drivers in the estimations. Although we expected an incentive effect deriving from the contract characteristics (chapter five of this thesis, Bradley, Green, Leeves 2012), the baseline results vary only marginally in coefficients. Thus, we are very confident that our findings are robust.

Still, our study entails some limitations as a consequence of its organizational setting. First, we do not have detailed information neither on the algorithm that determines performance evaluations nor on the underlying mechanisms that determine the total amount of incentives. It would have been interesting to work on the incentive scheme in more detail, e.g. by observing only those drivers at the verge to a higher incentive level. Second, the incentives were introduced for the first three month without informing the drivers. Still, there is the possibility that the information might be spread premature. Since we expect the results for this period to be biased due to information leaks, we exclude the respective three months window from our baseline estimations. Yet, checking for the introduction of incentives after three months supports our overall findings (see Table A.8 in the appendix). Third, the bonus payments incentivize low fuel consumption only indirectly by rewarding an eco-friendly driving style. Therefore, one might question the validity of our findings on the link between incentive scheme and fuel consumption. However, all target values of the driving parameters and evaluation scores set by the company aim at reaching eco-friendly driving behavior that yields at low fuel-consumption. Moreover, our findings on fuel consumption are entirely supported by estimations for raw performance indicators such as driving parameters and evaluation scores. Thus, we are confident to address doubts on the indirect relationship of incentives and fuel consumption. Finally, our data set includes only three women. Although there is support for a gender effect in self-selection and performance under incentives (e.g. Croson, Gneezy 2009, Dohmen, Falk 2011), we avoid working on this issue to account for the unequal gender distribution among drivers. Yet, the low share of women is industry specific and, therefore, should not bias or impact our overall findings.

4.5 Conclusion

With this study we offer new insights to the ongoing debate on the benefits and drawbacks of extrinsic incentives. While classic economic theory assumes employees to react to extrinsic (monetary) incentives by increasing performance and productivity, behavioral incentive theory suggests extrinsic rewards to crowd out other important intrinsic and social incentives that motivate people. There is broad experimental and empirical support for both theoretical assumptions.

To contribute to this discussion, we use a hitherto unavailable data set on the performance of commercial truck drivers collected from the internal fleet management system of the inhouse hauler of a large European truck manufacturer. Truck drivers receive bonus payments for good performance within the first twenty-four months of the thirty-six months observation period. Subsequently, management decided to abolish the incentives. We empirically measure the effect of the abolition of the incentive scheme on driver effort choice decisions by analyzing four separate sets of dependent variables of driver performance: fuel consumption, driving parameters, driving scores and trip evaluation. Consistent with behavioral incentive theory, we find comprehensive support for crowding-out effects of the bonus pay scheme applied by the hauler. In other words, we observe drivers to display significantly lower effort and performance when extrinsically incentivized compared to the non-incentive environment. These findings are in line with previous research (e.g. Deci 1971, Gneezy and Rustichini 2000a). More precisely, fuel consumption is significantly higher when incentives are in practice, adding up to additional 530 liters used per driver and year which translates to extra fuel costs of €800 per driver and year. We interpret this amount as hidden costs of incentives. Combined with the net bonus payments, the hidden costs constitute a serious financial burden to the company's baseline results – especially in a business as competitive as the trucking industry. When narrowing the observation period around the lagged information on the introduction and the final abolition of the bonus pay, we observe strong initial effects for both events. Further analyses of driving parameters suggest that drivers put less effort in adapting an eco-friendly driving style. This results in significantly lower values on driving scores: nine percentage points worse on hill driving, thirteen percentage points worse on anticipation behavior and even sixteen percentage points worse on use of brake. Concerning trip evaluation data, we observe a more than 70% higher probability of a trip being rated as "good" subsequent to the abolition of incentives. In short, our results indicate that drivers respond to incentives not the way intended by the hauler. Instead, it is the abolition of the incentive system that significantly increases driver performance. These findings remain robust on all four performance measures under various specifications and robustness checks. The results clearly support crowding-out effects of extrinsic rewards as stated by behavioral incentive theory. In this sense, the empirical evidence is in clear contrast to classical economic arguments on the benefits of monetary incentives (e.g. Gibbons 1998, Oyer, Schaefer 2011).

We believe our findings to arise from monetary incentives crowding out other important intrinsic and social motives that seem to play a major role in truck driver motivation. In other words, drivers are for some reasons intrinsically motivated to choose a certain level of effort. Though, an extrinsic incentive – such as a financial reward – seems to diminish or even disperse their intrinsic motivation. As a consequence, overall net effort choice and performance may be lower than without receiving financial rewards. There are some intrinsic and social motives that might serve as interpretation at this point. First, research identified individual environmental awareness and concerns to be essential for adapting pro-environmental behavior (e.g. Eagly, Kulesa 1997, Fransson, Gärling 1999). Therefore, we assume drivers' environmental beliefs to be a major intrinsic motive to adapt eco-friendly driving.[57] In line with this argument, extrinsic incentives are observed to crowd out pro-environmental behavior (e.g. Thøgersen 1994). Second, prior research has identified social image, status and reputation as powerful intrinsic incentives to people's motiva-

[57] For an overview on pro-environmental behavior of truck drivers see Schweitzer, Brodrick and Spivey (2008).

tion (e.g. Ariely, Bracha, Meier 2009, Ellingsen, Johannesson 2008). Hence, we assume drivers to have some social concerns on having an eco-friendly image and reputation, in particular as they are working in a high-polluting industry. This might be even strengthened as the country domestic to the drivers is far below OECD average on CO_2 emissions (OECD 2014b). Third, peer competition among drivers may represent an intrinsic incentive, too. According to company staff, the first thing drivers do when returning from a round-trip is to check their traffic light evaluation relative to peer-group performance. Theory and empirical evidence suggest that peer-group competition functions as social incentive. For instance, effort choice is higher when relative performance rankings are reported (e.g. Charness, Masclet, Villeval 2013). We believe that monetary incentives – at least partially – bought off the game-like character of peer competition. The same holds true for classic peer effects that motivate drivers to good performance via social pressure (e.g. Falk, Ichino 2006, Mas, Moretti 2009). Eventually, introducing incentives in the first place might be interpreted by drivers as an act of distrust. According to Sliwka (2007), feeling distrust might turn a person into selfish behavior. This effect could even be intensified by the secret introduction of the incentive scheme that most likely was seen as a lack of trust by drivers. Acting selfish might reduce truck drivers' concerns about both the environment and the competitiveness of their employer. Instead, they might increase their own utility by putting less effort in economic driving behavior. Furthermore, our findings support prior evidence on people responding strongly to both well-structured and badly designed incentives (e.g. Roberts 2010). In particular, incentives that reward performance with an amount too small may counteract the intended behavior (e.g. Gneezy, Rustichini 2000a, 2000b). By offering less than 5% of the monthly net income, the hauler might have badly designed the incentive by choosing an amount below the incentivizing threshold. This especially holds true given that drivers are working in a high-paying company in a country with a high wage level and, therefore, may anecdotally be ranked among the top-earning truck drivers in Europe.

However, translating our findings into clear implications to practitioners proves to be difficult. It seems that principals need to abandon the idea that extrinsic incentives serve as a panacea to motivational issues. Instead, monetary incentives do not work as intended per se and employee performance and productivity depend on other intrinsic and social motives, too. Introducing extrinsic incentives without seriously reflecting on these issues might prove detrimental to overall productivity in the end. Thus, we highly recommend considering the internal and external fit of incentive schemes and general HRM instruments. At this point, we follow Ichniowski, Shaw and Prennushi (1997) who argue that incentive plans work best when introduced with complementary HRM practices. Therefore, we like to encourage practitioners to carefully monitor incentive systems and to take appropriate actions if incentives appear to have counter-productive effects. Nevertheless, it seems to be good news that mistakes once made in incentive introduction or design seem reversible since employees quickly respond to new settings.

Eventually, the very particular organizational set up and the insider econometric approach of this paper might raise doubts on the generalizability of our findings. However, given that our results support theoretical assumptions stated by behavioral incentive theory and are in line with previous research, we are confident that our evidence is applicable to similar situations at least to some extent. Yet, we are not able to provide clear evidence on the generalizability of our findings to other organizational settings. We therefore encourage researchers to further contribute to the ongoing debate on both crowding-out and price effect of extrinsic incentives. Studies examining different organizational settings apart from insider econometrics might widen general insights on incentive effects on employees' effort choice decisions. In response to the so far inconsistent findings on extrinsic incentive effects on employee effort, research on moderating and mediating effects of incentives and the fit between intrinsic and extrinsic incentives seems to be highly appreciated.

5 Behavioral Consequences of the Transition from Temporary to Permanent Employment

5.1 Introduction

Temporary employment contracts have become a common practice in modern organizations to circumvent labor market regulations considered as restrictive and to flexibly react to changing staff requirements that are due to changes in product demand (e.g. Houseman 2001). Moreover, temporary employment offers firms opportunities to reduce labor costs as well as administrative complexity (De Cuyper et al. 2008). According to the most recent Eurostat Labor Force Survey, about 14% of all employment contracts in the EU-28 in 2013 were temporary in nature[58] (Eurostat 2014c). Among those on temporary contracts, younger workers are heavily overrepresented (OECD 2014c). In this paper we define temporary contracts to be different from standard contracts with respect to the following three dimensions: limited duration, various statutory drawbacks (no or little employment protection, no minimum wage etc.) and the different organizational setting (temporary contracts are often with temporary-work agencies (e.g. De Cuyper et al. 2008)).[59]

Temporary employment is not only used to avoid labor shortages and reduce labor costs but is often considered a “stepping stone” (Booth, Francesconi, Frank 2002) or a “port of entry” into permanent employment (Berton, Devincienti, Pacelli 2011, Buddlemeyer, Wooden 2011). Recent research emphasizes the nature of temporary employment as a screening device allowing employers to test the quality of a particular job match without the risk of long-term contractual obligations (e.g. Güell, Petrongolo 2001, Gagliarducci 2005). Moreover, a growing body of literature demonstrates that contract characteristics significantly affect employees’ behavior (e.g. Ichino, Riphahn 2005, Guadalupe 2003) in the sense that e.g. temporary and permanent workers are found to differ significantly in their choice of effort levels (e.g. Engellandt, Riphahn 2005). Although permanently employed workers are supposed to perform better than temporary workers because they receive more (firm-specific) training and report higher work satisfaction, empirical studies found the opposite to be true by demonstrating that temporary workers sometimes outper-

[58] In the country where the headquarters of the hauling company is located, the percentage of temporary employees is slightly above EU-28 average (OECD 2014).

[59] In this study, the term “temporary employment” refers to all kinds of contingent, fixed-term, casual or non-permanent employment relations with an employer characterized by an ex ante defined limited contract duration.

form permanent workers (e.g. Livanos, Zangelidis 2013, Amuedo-Dorantes 2002). We believe the reasons to be twofold. On the one hand, employers very often use temporary employment as a screening device. On the other hand, employees consider motivation and effort as strong signals to an employer that are likely to increase their probability of being promoted to a permanent contract. Both effects incentivize temporary workers to outperform permanent workers in terms of effort levels. However, this holds true only as long as temporary work is associated with inferior job characteristics compared to permanent employment. So far, temporary positions have been found to be associated with less favorable working conditions (e.g. Paoli, Merllié 2001), lower wages (e.g. Mertens, Gash, McGinnity 2007), less training (e.g. Arulampalam, Bryan, Booth 2004) and lower levels of job satisfaction (e.g. Boyce et al. 2007). Moreover, the uncertainty about future job prospects can be a mental burden to temporary workers resulting in health problems (Sverke, Hellgren, Näswall 2002). Finally, temporary workers do not benefit from mandatory job protection legislation to the same extent as permanent workers. Given these disadvantages, temporary workers should have strong incentives to demonstrate motivation and effort to “qualify” for a promotion from a temporary to a permanent contract.

These considerations inevitably lead to the question of whether and to what extent formerly temporary workers adjust their effort level after having signed a permanent contract. One might expect significant behavioral responses from employees who have been offered more favorable working conditions, including dismissal protection legislation. Using an unbalanced panel of truck drivers from the GPS-based fleet management system of an in-house hauler of a large European truck manufacturer we are among the first to address this issue with individual productivity information from employees whose contract status was changed from temporary to permanent. Our results confirm a significant reduction in commercial truck drivers’ effort levels resulting in higher fuel consumption and a poorer overall driving performance after being promoted from temporary to permanent employment.

Our study contributes to the literature in several regards. First, we add to the growing research on the influence of employment contract characteristics on employee behavior. In particular, we extend the literature on workers’ effort choices when contract characteristics become more employee-friendly (e.g. Ichino, Riphahn 2005). Second, we are among the first to use data from a modern GPS-based fleet management system to analyze driver behavior. This usually extensive source of data has not yet made its way into the personnel economics literature. Third, we follow the insider econometrics tradition, which has enjoyed increasing popularity among personnel economists since the publication of the seminal papers by Ichniowski, Shaw and Prennushi (1997) on HRM practices in U.S. steel plants and Lazear (2000b) on monetary incentives for windshield installers. The insider econometrics approach applies sophisticated econometric methods to a (panel) data set gathered in one or a few companies. The resulting nano-perspective allows detailed insights into the behavior of individuals upon changes in incentives, working conditions, etc.

The remainder of this paper is structured as follows. The next section provides both a review of the literature on the incentive effects of temporary employment as well as on employees' behavioral responses to changes in contract characteristics. The specific organizational setting and the data set are described in detail in section three. Section four presents the estimation strategy as well as the results while section five concludes.

5.2 Literature Review and Theoretical Framework

It is a stylized fact in the personnel economics literature that temporary workers receive less training by employers due to the usually rather short duration of the employment relation (e.g. Arulampalam, Bryan, Booth 2004, Hoque, Kirkpatrick 2003, Forrier, Sels 2003, Arulampalam, Booth 1998). As a consequence, temporary workers accumulate less firm-specific knowledge than permanently employed workers. Since firm-specific skills are critical for performance (e.g. Hatch, Dyer 2004, Hitt et al. 2001) permanently employed workers are expected to outperform temporary workers.

However, there is also evidence showing that despite their lower levels of training temporary workers often outperform employees with a permanent contract. The performance differential is likely due to temporary workers' higher levels of motivation and work effort. Using the Swiss Labor Force Survey Engellandt and Riphahn (2005), for example, find that the probability of working unpaid overtime is 60% higher for temporary compared to permanent workers. Moreover, in a large data set from Australia Bradley, Green and Leeves (2007) find evidence of significantly lower levels of absenteeism (their preferred measure of employee effort) among temporary workers. This latter result is confirmed by Livanos and Zangelidis (2013) for the EU, Arai and Thoursie (2005) for Sweden and Amuedo-Dorantes (2002) for Spain. Moreover, using data from a Spanish survey Guadalupe (2003) finds that other things equal temporary workers have a five percentage points higher accident probability which may be due to lower investments in temporary workers' human capital as well as their pronounced incentives to demonstrate high levels of motivation and effort in the sense of taking higher risks.

The incentive effects of temporary contracts that are assumed to result in higher motivation and effort levels of temporary employees are due to their specific screening and signaling properties (Spence 1973). First, temporary contracts are often used as a screening device to test newly hired workers before they are being offered a permanent contract (e.g. Green, Leeves 2004). Second, temporary work may be considered a "stepping stone" into permanent employment (Booth et al. 2002). Consistent with the latter argument Güell and Petrongolo (2001) find that temporary workers often sign a permanent contract before their initial temporary contract had expired. This suggests that firms use temporary employment as a screening opportunity to discover workers' "true" qualities and offer permanent contracts as soon as the required abilities have been identified (see also Buddlemeyer and Wooden (2011) with comparable evidence for Australia and Gagliarducci (2005) for Italy). Summarizing, these studies suggest that temporary contracts are used as a screening instrument that enables employers to reduce information asymmetries about workers' motivations and effort choices without the risk of long-term contractual obligations.

Summarizing, the characteristics of temporary contracts should provide strong incentives for temporary workers to choose high effort levels to produce those signals that maximize the probability of being hired on a permanent contract. This holds true, however, only as long as permanent employment proves to be advantageous over temporary employment. Indeed, temporary employment is usually associated with poorer working conditions and lower pay levels for equal performance. On the one hand, temporary workers have been found to benefit less from investments in health care or ergonomics leading to higher levels of job dissatisfaction, fatigue and muscular pain (Benavides, Benach 1999, Benavides et al. 2000). On the other hand, they are exposed more often to repetitive tasks and movements as well as heavy loads (Paoli, Merllié 2001), enjoy lower levels of work autonomy and report themselves the least well-informed about their work environment (Aronsson 1999). Furthermore, temporary workers often consider themselves subject to "status stigmatization" leading to lower levels of well-being, job satisfaction, and performance (Boyce et al. 2007). Generally speaking, temporary jobs are found to be of an overall lower job quality (Green, Kler, Leeves 2010). The wage penalty of temporary workers is well documented in the literature: Mertens, Gash, McGinnity (2007) for example find that the wages of temporary employees are 6.0% lower in Germany and 4.4% lower in Spain compared to employees with permanent contracts. Holmlund and Storrie (2002) find a wage penalty of around 10% in Sweden. Thus, employees on temporary contracts should be highly motivated to compete for promotion into permanent contracts. Reinforcing these incentives is that job insecurity has been found to be detrimental to health and well-being, leading to increasing psychological morbidity (Virtanen et al. 2005) and overall job stress (De Witte 2005)[60].

In addition, strict employment protection for workers on permanent contracts is often associated with low effort levels (e.g. Riphahn 2004, Ichino, Riphahn 2004) as employees feel sheltered by law from employer actions. Temporary workers who are usually exempt from labor protection legislation should, first, exert more effort to avoid negative consequences, such as a lay-off. Second, temporary workers have strong incentives to exert high levels of effort to recommend themselves for permanent contracts that, in turn, are associated with stricter employment protection. Summarizing, temporary contracts are associated with less favorable characteristics than permanent contracts. Thus, temporary workers have strong incentives to display high levels of effort in case that employers use temporary contracts as a screening device. High effort signals increase the probability of qualifying for tenure, leading to better working conditions, higher pay and stricter employment protection (Amilon, Wallette 2009).

In this paper we seek to answer the question whether and to what extent employees' effort choices are affected by changes in contract status, i.e. by being promoted from a temporary to a permanent contract. From an economic point of view (individuals are assumed to be utility maximizers) a change is to be expected: As soon as the incentivizing effect of a temporary contract is gone, a formerly temporary worker is not motivated any longer to choose a high effort level. This assumption has been documented in a body of literature studying the behavioral responses of employees after the end of their probation period (during probation, i.e. in the first three to six months in a new job, employees do not benefit from mandatory employment protection legislation and can, therefore, be dismissed easily). Thus, this instrument is popular among employers to screen the qualities of new recruits without having to accept long-term contractual obligations from the onset. Similar to temporary employees, workers during their probation period are assumed to have strong

[60] Two recent meta-analyses by Sverke, Hellgren and Näswall (2002) and Cheng and Chan (2008) summarize the health-related consequences of job insecurity.

incentives to signal that they are able to meet and exceed the employer's expectations in order to increase the probability of staying with the company and qualifying for employment protection. There is growing evidence in the literature confirming a behavioral response of employees reducing effort choice and increasing shirking after the probation period ends. Using the "German Socio Economic Panel", a large and representative sample of the German working population, Riphahn and Thalmaier (2001) find evidence of this behavioral (moral hazard) effect by demonstrating a statistically significant increase in absenteeism as soon as legal employment protection sets in after probation. Using a large sample of employees working for an Italian bank with branches all over the country, Ichino and Riphahn (2005) also find a significant increase in individual absenteeism once the employees' probation periods have come to an end and they are sheltered by mandatory employment protection laws. Finally, Pfeifer (2010) using personnel data from a large German company finds that newly hired white-collar workers are more than 50% less likely to be absent during their probation period than in the following nine months.

Given the behavioral responses of employees after probation, this argument has recently been adapted to explain changes in the behavior of (formerly) temporary workers who have been signed on permanent contracts. Using a representative sample of Australian employees, Bradley, Green and Leeves (2012) report significantly higher levels of effort (in terms of lower absence rates) for temporary workers who are assumed to be incentivized by being rewarded with a permanent contract. They find a statistically significant increase in individual absenteeism for workers moving from temporary to permanent employment. Summarizing, the available evidence leads us to conclude that temporary workers have strong incentives to reduce their effort levels after having signed permanent contracts as they do no longer need to demonstrate high levels of effort or avoid behaviors that might be considered by employers as shirking. Thus, we expect employee performance to deteriorate after a change in contract status (from temporary to permanent).

We extend the available literature by applying the concept of behavioral changes following a change in contract status, i.e. when employees are promoted from a temporary to a permanent job. We are, to the best of our knowledge, the first researchers to work on this question with individual performance data other than absenteeism. Our data set from the trucking industry offers two unique and perfectly objective performance measures – individual fuel consumption as well as a performance evaluation automatically generated by the company's GPS fleet management system. Moreover, we add to the insider econometrics literature by using a unique data set from one particular company. Finally, by using data derived from the fleet management system of the in-house hauler of a large European truck manufacturer we are, again to the best of our knowledge, the first economists to literally open the treasure box of this promising data source for personnel economics research.

Investigating the issue of behavioral responses to changing contract characteristics in the trucking industry is promising in three ways. First, driver performance in terms of driving style immediately influences costs and thus operational efficiency. Since trucking is a highly competitive industry, even small differences in drivers' effort choices can significantly affect a company's profitability. Therefore, research resulting in implications for firms to increase their operational efficiency should be appreciated by both, academics and practitioners. Second, while in the past the carrier-driver relationship was considered a prototypical principal-agent setting with the agent performing his tasks beyond the principal's monitoring range,[61] modern GPS-based on-board electronic computer systems now make driver monitoring an easy task (Hubbard 2000, Baker, Hubbard 2004). While until recently only the final result of a commercial truck driver's performance could be measured (amount of goods transported and arrival time), modern fleet management systems now allow for an evaluation of the way the trip was carried out by monitoring driver behavior while on the road (Hubbard 2000) providing haulers via on-board computers with reliable and real-time data on driver performance. Thus, these systems offer indisputable evidence of driver performance and, at the same time, discourage drivers from manipulation. Third, the performance of individuals is unaffected by the performance of their peers, suggesting that team effects arising in other environment are completely absent here. However, the performance of commercial truck drivers remains subject to exogenous conditions the drivers themselves cannot influence, e.g. weather and road conditions or overall traffic. All this can influence driver performance in ways that conclusions drawn from the electronically assembled performance data may be misleading. However, our data set enables us to control for these external effects. These controls as well as the data set will be discussed in more detail in the following section.

[61] For a first application of the principal-agent model in the trucking industry see Vernon and Meier (2012).

5.3 Data Set and Descriptive Statistics

To analyze employees' behavioral responses to a change in contract status, we use a unique panel data set generated from the internal GPS-based fleet management system of an in-house hauler of a large European truck manufacturer. The data covers 36 consecutive months (from January 2011 to December 2013) and has detailed information on commercial truck drivers' performance on a trip-by-trip basis. The data was recorded via on-board computers installed in every truck and includes information on various driving parameters such as fuel consumption, distance covered and a large number of truck handling variables (e.g. average speed, speeding, number of brake applications, etc.). Moreover, the on-board computers assess driving performance by means of an internal traffic light evaluation tool. The data is supplemented with external information we believe to influence traffic conditions and truck handling (e.g. weather, time of the day, day of the week, etc.). Additionally, the data set was expanded to include information from the drivers' personnel records (e.g. age, tenure and current contract status, i.e. temporary or permanent).

The drivers' assignment is to deliver intermediate products from one production facility of the truck manufacturer to another by passing through four Northern European countries, covering a distance of 1,250 kilometers. Since drivers are legally required to strictly comply with mandatory driving and resting periods[62], the hauler operates a resting facility approximately halfway. There, sleeping, cooking and sporting facilities are provided for use during the mandatory break. A typical route for a single truck is as follows: Driver A arrives and immediately hands over his truck to driver B, who has just finished his resting time. Driver B continues the trip to the final destination making sure that the goods arrive just-in-time. At the same time, driver A awaits the arrival of the next truck. After having finished his mandatory rest period driver A then continues his journey to the final destination on a truck that has just been handed over to him. Thus, on a standard round-trip each driver completes four different stages on four different vehicles with the drivers having no influence whatsoever on the kind of truck (the trucks are different in terms of age and in terms of brands).

The initial data set includes information on 81 drivers (n=11,439 driver-trip-observations). However, we exclude observations of drivers whose contract status did not change during

[62] According to regulations on road haulage operations by the European Commission the daily driving period must not exceed 9 hours while daily resting periods of at least 11 hours are mandatory. Drivers may slightly deviate from these regulations two or three times a week (European Commission 2014d).

the observation period resulting in n=2,769 driver-trip-observations. We further restrict the dataset to include only trips with a reported distance between 520 and 720 kilometers since the production sites are located 570 kilometers and 670 kilometers away from the resting facility[63]. Moreover, we only retain in our data set drivers with at least 30 trips before and 30 trips after a change in contract status. Finally, we exclude trips with "suspicious" levels of fuel consumption, i.e. a reported fuel use outside a range of minus two standard deviations to plus four standard deviations from the mean.

These restrictions lead to a small sample including eight drivers only (these drivers represent 10% of the company's workforce in long-haul trucking).[64] More important, however, is the fact that we have at least 100 observations for each of the eight drivers (mean=162.4 observations). The final number of driver-trip-observations (n=1,299) appears to be sufficient for a detailed econometric analysis as the data set covers approximately 11.4% of all trips during the period under investigation.

As already mentioned above, the information from the fleet management system was matched with driver demographics including age and tenure as well as information on the exact date when a driver's contract was converted from temporary to permanent (Table 5.1 displays the summary statistics of drivers' demographic characteristics.

Table 5.1: Summary Statistics – Driver Demographics

Variable	***mean***	***s.d.***	***min.***	***max.***
Driver age (in years)	42.4	12.3	19	61
Driver tenure (in 3-months intervals)	5.9	3.3	1	13
Driver temporary status (temporary status=1)	.44	-	0	1

Number of drivers: 8
Number of driver-trip observations: 1,299

Interestingly, reported driver age covers almost the complete working life ranging from 19 to 61 years with a mean of 42.4 years (s.d.=12.3). Comparing the age of these eight drivers

63 Distances shorter than 570 km occur if drivers exceed the maximum driving time allowed before reaching either the resting facility or one of the production sites (due to e.g. traffic jams). Distances longer than 670 km arise if drivers have to avoid traffic congestions or constructions by taking alternative routes.

64 In addition to long-distance trucking the hauler operates domestic short-haul trucking as well as further transportation services (e.g. bus transportation, etc.).

with that of the remaining 73 trucker, we see that the figures are almost identical (mean=44.2, s.d.=11.0, min=22, max=64). The same holds true with respect to nationality as all drivers are reported to originally come from the hauler's domestic country. With respect to tenure – reported in three-month intervals – we find slightly lower values for the eight drivers who were initially employed on a temporary contract. Of the 1,299 driver-trip observations that we use in our estimations, 44% occurred when the drivers were on a temporary contract while 56% occurred after the drivers' contracts had been converted to permanent. This allows for a "clean" pre-post contract status comparison.

The hauler operates a modern GPS-based fleet management system to manage and monitor all activities of its fleet. Each vehicle is equipped with on-board computers providing abundant information on fuel consumption, driving style and truck handling. Furthermore, the system automatically assesses driver performance along selected driving parameters. With the help of a complex algorithm – which is unknown to us – the on-board devices compute evaluation scores based on target values predefined by the hauler. Four of the most important evaluation criteria are anticipation, use of brakes, hill drive and choice of gear. These scores are then merged with further driving parameters to an overall trip evaluation classifying each trip as either "good", "mediocre" or "poor". Based on a traffic light system, the evaluation is reported to an online tool which allows fleet managers to monitor each driver's performance on every single trip. Interestingly, drivers can also access these evaluation reports and compare their results with the performance of their peers.

Table 5.2: Summary Statistics – Dependent Variables

Variable	***mean***	***s.d.***	***min.***	***max.***
Fuel consumption				
Fuel consumption (in l/100km)	28.0	2.9	21.6	38.7
Trip Evaluation				
Overall (good=1, med=2, poor=3)*	1.89	.98	1	3
Good (yes=1)*	.54	-	0	1
Mediocre (yes=1)*	.03	-	0	1
Poor (yes=1)*	.43	-	0	1

Number of drivers: 8
Number of driver-trip observations: 1,299

* Reduced number of observations due to three missing values (n=1,296).

It is important to bear in mind that the traditionally used measure “fuel consumption per 100 km” is not included in the automatic evaluation of the individual trips. However, driving behavior that leads to a “good” evaluation is characterized by an eco-friendly driving style (no harsh accelerations, no speeding, etc.). It appears from Figure A.1 in the Appendix, that fuel consumption and trip evaluation are uncorrelated: Average fuel use for trips evaluated as “good” is 28.1 liters per 100 km, for trips evaluated as “mediocre” it is 28.2 liters and for trips evaluated as “poor” the respective value is 27.9 liters ($F=0,49$; not significant). Moreover, it appears from Table 5.2 that slightly more than half of all trips (54%) are assessed as “good” and 43% as “poor”. Only 3% of all trips are evaluated as “mediocre”. Thus, apart from the standard performance measure “fuel use per 100 km” we use in our estimations a second performance measure (“quality of trip as assessed by the company’s fleet management system”) that is uncorrelated with the first one.

Our study design has a number of advantages avoiding major weaknesses of previous research in the economics of trucking. First, drivers can usually not be compared as they work under highly variable conditions, experiencing different exogenous effects of different routes and destinations (e.g. road conditions, constructions, detours, etc.). Due to the organizational setting of our study we are able to assume equal road and environmental conditions since all drivers use the same predefined route. Second, based on detailed information on date and time of each trip, we are able to control for weather as well as for date- (e.g. holiday) or time-of-the-day related (e.g. rush hours) peaks in traffic density since “timing” is likely to influence driving performance. Third, drivers are observed on different trucks without having any influence on truck selection. Thus, we can account for vehicle specific effects and eliminate any impact of driver-truck matches. Fourth, almost one quarter of all vehicle-kilometers of trucks in European cross-border traffic occurs on empty trucks (European Commission 2014c). Any study that is unable to control for load factors is thus likely to suffer from an omitted variable bias as full and empty trucks differ in maneuverability and fuel consumption. Since the hauler whose data we use here reports average load factors of more than 90% on its out as well as its return journeys, this issue can be neglected.

5.4 Empirical Models and Estimation Results

Recall that our empirical analyses are based on two different, yet somewhat related performance measures that can be considered proxies for choice of effort. First, we use in our estimations average fuel consumption per 100 km as the dependent variable. It appears from Table 5.2 that mean fuel consumption is 28.0 liters/100km with a standard deviation of 2.9 liters. Since all drivers receive individual fuel efficiency training in irregular intervals, we consider them well-informed on how driving behavior and style affects fuel consumption.[65] Thus, we suggest any differences in fuel consumption to derive from drivers' individual effort choices. Second, we use the reports of the internal trip evaluation tool assessing each trip as either "good", "mediocre" or "poor" (see Table 5.2). Since drivers are offered full online access to all evaluation reports they are aware not only of their performance relative to their peers, but also of any kind of "misconduct" in driving behavior. Interestingly, publicly available performance evaluations have recently been found to eliminate social loafing and increase performance at least in group work settings (Lount, Wilk 2014). Further research indicates feedback to be highly appreciated by truck drivers with feedback from advanced technology being less desired than feedback from supervisors (Huang et al. 2005). Thus, drivers receiving performance feedback and still not performing in accordance with the employer's expectations are assumed to drive with reduced effort. While fuel consumption is a categorical variable, the trip evaluation variable is ordinal. Both outcome variables are computed by the in-vehicle computers and are not prone to any subjective rater bias or manipulation by the drivers.

Our models have the following general form (with *FUELC* denoting fuel consumption per 100km and *EVAL* being the composite evaluation measure):

$$FUELC = \alpha + \beta DRIV + \gamma VEH + \delta TRAF + \varphi WEA + \omega CONTR + \varepsilon \qquad (1)$$

$$EVAL = \alpha + \beta DRIV + \gamma VEH + \delta TRAF + \varphi WEA + \omega CONTR + \varepsilon \qquad (2)$$

where α is the constant, *DRIV* represents the driver performance (see equation (3) for details), *VEH* is a set of vehicle dummies, *TRAF* a vector of traffic conditions, *WEA* a vector of weather conditions *and CONTR* a vector of further controls, β, γ, δ, φ and ω are the es-

[65] Although we have no information on the frequency or the exact dates of individual fuel efficiency training we assume this not to be a problem since recent research has convincingly demonstrated that (car) drivers seem to know by intuition how to drive eco-friendly (van Mierlo et al. 2004). Therefore, we assume that fuel efficiency training does not fundamentally change driver behavior, i.e. fuel consumption goes down after training but returns to its pre-training level rather quickly.

timated coefficients and ε is the error term. All variables are discussed in more detail below.

Driver performance (*DRIV)* is modeled as follows:

$$DRIV = f\,((AGE + HUMC) * EFCH) \qquad (3)$$

where *AGE* is driver age and *HUMC* represents a drivers' human capital in the form of past driving experience with the current employer (tenure in three-month intervals).[66] Previous research provides evidence for the negative effect of age on various abilities relevant to driving, e.g. response time, visual and psychomotor abilities (e.g. Llaneras et al. 1998). Other findings, however, suggest that older drivers can compensate for age-related impairments by experience (e.g. Guest, Boggess, Duke 2014). Additionally, young and less experienced drivers are found to have a higher probability of being involved in accidents (e.g. Häkkänen, Summala 2001), which further underlines the importance of experience in trucking. Hence, we include age as well as driving experience in our model to account for these effects. Both variables taken together represent the performance potential of an employee, which is then subject to drivers' current effort choice (*EFCH*). Depending on the choice of effort, performance potential eventually translates into actual driver performance. In our estimations *EFCH* is represented by a dummy variable taking a value of 1 of a driver is on a temporary contract and 0 if he is on a permanent contract.

In addition, we control in our estimations for exogenous effects that may influence truck driver performance. *VEH* is a series of vehicle dummies accounting for differences in technological configurations of trucks that might facilitate driving and/or decrease fuel consumption. *TRAF* is a vector of variables controlling for traffic density on the trips in two different ways. On the one hand, we include variables to control for the time of the day when the streets are presumably less congested (e.g. during night time, between 8 p.m. and 7 a.m. the next day) or more congested (e.g. during rush hours, between 7 a.m. and 9 a.m. in the morning as well as between 4 p.m. and 6 p.m. in the afternoon) when traffic density is particularly high due to commuter traffic. On the other hand, we include the day of the week as we believe some days to be systematically different (e.g. Mondays or Fridays) from others. Furthermore, we control for a three-day window around international

[66] We are well aware that tenure on a monthly basis would be preferable for the first few months; however, we chose three-month intervals because during our observation period of 36 months any effect of monthly tenure is likely to disappear quickly as time elapses.

holidays during which we suppose traffic density to be higher due to holiday traffic (public holidays in all countries the drivers pass on their trips, e.g. New Year's Day, Easter, Christmas, etc.). *WEA* is a vector of weather dummies including wind speed, precipitation, snow depth, and temperature[67] because previous research has found driving behavior, traffic conditions as well as crash risks to be associated with weather conditions (e.g. Kilpeläinen, Summala 2007). Moreover, we include dummy variables for all thirty-six months to account for possible trend or timing effects. Finally, to control for season effects we include dummies for all fifty-two weeks of a particular year. Given the wide range of control variables, we are confident to have excluded most exogenous influences on driver performance and are thus able to study the "clean" effect of drivers' behavioral responses to a change in their contract status.

We start the discussion of our findings by looking first at the impact of contract status on fuel consumption (equation (1) above) in the short-term, followed by a detailed analysis of its impact in the long-run. We then go on to present our findings with respect to the impact of contract status on trip evaluation (equation (2) above). The main results are reported in Tables 5.3, 5.4, and 5.5 suggesting that commercial truck drivers significantly change their behavior following changes in their contract status, i.e. reduce their effort levels after having been promoted from a temporary to a permanent contract. The baseline results of our fixed-effects estimates with fuel consumption as the dependent variable are displayed in Tables 5.3 and 5.4. We use various specifications including different sets of explanatory variables to control for the robustness of our findings.

To identify the impact of a change in contract status on fuel consumption in the short-run, we narrow the observation period to a six-month window and compare for each driver his performance in the last three months under the temporary contract with that during the first three months under the permanent contract. It appears from Table 5.3 that the immediate consequence of promoting a driver from a temporary to a permanent contract is a statistically significant and economically relevant increase in fuel consumption of more than 10% (3.04 l/100km). While this finding supports our assumption of a strong initial behavioral response in the first few months after having been promoted to a permanent contract, it is

[67] Weather information was collected from a meteorological station located approximately in the middle of the round-trip route. This admittedly crude weather proxy helps to overcome the problem that we cannot control for truck position in detail. Thus, we assume that weather information from half the distance adequately reflects the slightly differing climate zones of the trips' starting and arrival points. Data was provided by the German Meteorological Service (DWD, 2014).

not yet clear whether this initial reaction remains constant over time or whether drivers change their behavior once more after some time has elapsed.

Table 5.3: Impact of Contract Status on Short-Term Fuel Consumption

Variables	**(1)**	**(2)**	**(3)**
Temporary Contract (yes=1)	-1.887**	-1.887**	-3.035**
	(.628)	(.628)	(.903)
Age (in years)		.480	1.079
		(4.215)	(4.188)
Tenure (in three-month intervals)			-1.445
			(.872)
Vehicle Dummies	YES	YES	YES
Date Dummies	YES	YES	YES
Time Dummies	YES	YES	YES
Weather Dummies	YES	YES	YES
Driver Fixed-Effects	YES	YES	YES
Constant	37.31	25.46	-1.461
	(72.72)	(170.0)	(171.5)
Number of observations	308	308	308
Number of drivers	8	8	8
R-squared	.343	.343	.350

Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.10

Table 5.4 displays the results of our second estimation, identifying the long-run behavioral responses of commercial truck drivers to changes in contract status. According to the point estimates, drivers on average reduce their individual effort level upon a change in contract (e.g. from temporary to permanent). The coefficient of contract status in our preferred specification (model 3) is -.475 (standard error = .150). This implies that, other things equal, a driver uses .475 l/100km less fuel when employed on a temporary contract. On average, drivers cover an annual distance of 125,765 km and use 35,185 liters of fuel. The estimated coefficient translates into an increased fuel use of 600 liters per year after a change in contract status. Assuming a fuel price of €1.50 per liter, this amounts to additional fuel costs of around €900 per driver and year for everybody who is promoted from a temporary to a permanent job. Thus, these costs can be interpreted as “hidden costs of contract conversion” (from temporary to permanent).

Table 5.4: Impact of Contract Status on Long-Term Fuel Consumption[68]

Variables	**FE**				**RE**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Temporary Contract (yes=1)	-.397* (.180)	-.415** (.161)	-.475** (.142)	-.487** (.150)	-.487*** (.151)
Tenure (in three-month intervals)				-.0740 (.208)	-.0026 (.209)
Age (in years)				-.604 (1.550)	-.0740 (.027)
Date Dummies	YES	YES	YES	YES	YES
Vehicle Dummies		YES	YES	YES	YES
Time Dummies			YES	YES	YES
Weather Dummies			YES	YES	YES
Driver Fixed-Effects	YES	YES	YES	YES	
Constant	25.51*** (1.085)	25.65*** (1.126)	.118 (25.37)	39.11 (64.72)	0 (0)
Number of observations	1,299	1,299	1,299	1,299	1,299
Number of drivers	8	8	8	8	8
R-squared	.090	.116	.128	.128	

Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.1

To further control for the influence of age and tenure we estimated a random-effects model, the results of which are displayed in column 5 of Table 5.4. Contrary to previous research (e.g. Llaneras et al. 1998), we find driver performance to be unaffected by age and/or experience. This is perhaps surprising as the drivers in our data set are very heterogeneous in terms of age (19 to 61 years) and tenure (up to 38 months). The fact that a driver's human capital (in terms of experience) seems to be irrelevant for his style of driving renders the impact of motivation and choice of effort (*EFCH*) even more important. Perhaps surprisingly, none of the thirty-eight vehicle dummies comes close to statistical significance (coefficients not displayed in table to save space). This finding might suggest that vehicle technology is not as important as driver behavior in determining fuel consumption. Yet, we believe this finding to be due to a more straightforward explanation. The hauler sorts out vehicles from its fleet after they have been used for about one million kilometers. Since this threshold level is usually reached after about two to three years, the hauler's fleet consists of almost (brand) new trucks and all drivers are equipped with state-of-the-art technology. Therefore, it seems reasonable that our dummies representing differences in vehicle technology fail to reach statistical significance. Interestingly and in contrast to previous research (Kilpeläinen, Summala 2007), the only statistically significant

68 Age and tenure effects were estimated individually but fail to reach statistical significance.

weather variable in terms of fuel consumption is wind speed (+.11 l). This is again plausible as trucks are prone to windage due to their size and unfavorable aerodynamic features. None of the remaining weather variables, e.g. temperature, snow fall and precipitation came close to statistical significance. In particular, we expected high temperatures to have a noticeable influence on fuel consumption since air conditioning systems are widely known to increase fuel use. With respect to the time dummies (months and weeks), no clear pattern emerges. On Tuesdays fuel consumption levels are higher (plus .77 liters per 100 km) higher than on the reference day (Mondays). No other day of the week effect can be identified.[69]

The results of our fixed-effect estimations still hold after including a dummy variable representing presence of a bonus regime during the first two years of the observation period.[70] The bonus regime might be an additional factor influencing driver behavior via choice of effort (*EFCH*). In particular, the bonus regime might impact the motivational effects that we attribute exclusively to behavioral responses following the change in contract status. However, we fail to find any significant effect of the bonus regime (see Table A.11 in the appendix with estimations including the bonus regime). Thus, we are confident that the behavioral responses identified are entirely attributable to changes in motivation and effort choice triggered by a change in contract status.

Since we are particularly interested in the timing as well as the persistence of the individual drivers' behavioral response, we included in our estimations a trend variable counting the number of months after contract conversion for each driver. The respective coefficient indicates a statistically significant negative effect (-.140), suggesting an increase of fuel consumption after the change in contract status (see Table A.12 in the appendix). The coefficient of the squared time trend, however, is positive and significant (.004) indicating a u-shaped pattern (with the turning point after 18 months). This, in turn, suggests that truck drivers in the long run return to their initial performance levels. Apparently, utility-maximizing truck drivers seem to no longer feel the need of displaying high levels of effort

---

69 The coefficients for trips at night (-.008 l) and during evening rush hours (+.26 l) have the expected signs but fail to reach statistical significance. Surprisingly, the coefficient for morning rush hours (-.11 l) is negatively signed, but also fails to reach statistical significance.

70 The bonus regime was in practice from the beginning of the observation period in January 2011 until December 2012 when the hauler decided to abolish it. Although a low fuel consumption was not directly rewarded by the incentive system, drivers were financially incentivized to perform according to the measures of the on-board evaluation tool.

after having been signed to a permanent contract. Indeed, individuals reduce their effort levels substantially and performance deteriorates considerably once a transition in contract status has occurred. However, with respect to fuel consumption this change in behavior seems to get weaker over time.

We now present the findings of our second estimation using the automatically generated trip evaluation as a measure for driving performance and proxy for driver choice of effort. To account for the ordinal nature of the data, we estimate a series of ordered logit models as suggested by Gujarati and Porter (2009) as well as Wooldridge (2013). It appears from Table A.12 in the appendix that there are significant behavioral responses in terms of higher levels of effort depending on contract status (temporary vs. permanent). Table 5.5 presents the post-estimation probabilities for individual trip evaluations.

Table 5.5: Estimated Probabilities for Trip Evaluation

Variables	Temporary	Permanent	Change	95% CI for Change	
Trip Evaluation					
Good	.7471	.4140	-.3331	-.4837	-.1825
Mediocre	.0404	.0559	.0155	-.0045	.0355
Poor	.2124	.5300	.3176	.1682	.4671

According to our estimations the probability of a trip being evaluated as “good” is 74.7% when a driver is on a temporary contract and 41.4% when a driver is permanently employed. Thus, the probability of a performance that is in accordance with the employer’s expectations is 33.3 percentage points (or 80.4%) higher in the case of the former workers. An inverse pattern appears for trips evaluated as “poor”. Here the probability for temporary workers is 21.2% compared to 53.0% for drivers on permanent contracts. This yields a difference of 31.8 percentage points or 150%, respectively. This result confirms our main finding that drivers on temporary contracts have strong incentives to signal motivation and high levels of effort. After having signed a permanent contract, drivers quickly reduce their effort levels leading to a lower overall driving performance evaluation. However, a closer look at the data reveals that when comparing both fuel consumption and trip evaluation before and after promotion to a permanent contract, only half of the drivers perform worse after they have been promoted. Thus, while some drivers show (negative) behavioral reactions to a change in contract status, others do not (this finding is in line with e.g. Nagin et al. 2002 as well as Riphahn and Thalmeier 2001).

We admit that due to the organizational setting of our study we have to be aware of some limitations that preclude generalization of our findings. First, the number of drivers (n=8) is quite small. However, the data set consists of at least 100 observations per driver resulting in a total number of observations of 1,299. We are, therefore, confident that our results are robust and reliable. Second, due to the hauler's data privacy restrictions of drivers' personnel records, we are unable to include in our estimations measures of driver qualification and former experience with other employers. Yet, since we control for age and tenure with the current employer, we believe to have included sufficient proxies for qualification and past driving experience. Moreover, the coefficients of both, age and tenure failed to reach statistical significance indicating a negligible importance of qualification and (former) driving experience.

5.5 Conclusion

Using a hitherto unavailable data set compiled from the fleet management system of an in-house hauler of a large European truck manufacturer, we provide robust evidence on employees' behavioral responses to changes in contract status (i.e. following the transition from a temporary to a permanent contract). We analyze commercial truck drivers' effort choices before and after they have been signed to a permanent contract using two performance variables – fuel consumption per 100 km and an automatically generated trip evaluation measure.

In line with previous research (Bradley et al. 2012), our findings suggest that commercial truck drivers choose their effort levels depending on the nature of their contracts, i.e. they exert more effort (in the sense of using less fuel and better trip evaluations) when on temporary contracts. After having been promoted drivers adjust their effort levels, i.e. use more fuel and receive more "bad" evaluations. These findings are robust on various specifications and survive a number of robustness checks. Drivers on temporary contracts (or more general: workers) – being aware of or just assuming screening activities to be used by the hauler (or more general: firms) – choose high levels of effort to signal motivation and dedication. However, as soon as drivers (workers) are signed to permanent contracts, they reduce effort back to "normal" levels. When employed under a permanent contract, fuel consumption of drivers is 600 liters higher per year (the additional costs per driver are around 900 € per year) than when employed under a temporary contract. We interpret this as the hidden costs of changes in contract status (from temporary to permanent). This negative (expensive) behavioral response is particularly strong in the first months after contract conversion but then declines to reach the initial level again after about 18 months. Moreover, for drivers on temporary contracts the probability of a trip being evaluated by the trucks' computer system as "good" is 55% higher than for drivers on permanent contracts. Yet, we find only half of the drivers to display these behavioral changes when being promoted from a temporary to a permanent contract. The other half does not respond to the new contract status but instead perform at their initial effort levels. This finding is in line with Nagin et al. (2002) who also find that only a fraction of employees behave as "rational cheaters" when given the opportunity to do so while many others resist that temptation.

Where do the incentive effects of temporary contracts come from? Permanent contracts have been found to be associated with better working conditions, higher pay, more job security and higher levels of employment protection. Most temporary workers want to recommend themselves for permanent contracts to enjoy increased levels of job security and dismissal protection legislation. This is particularly true in our case as the country where the headquarters of the hauler is located has most recently been ranked below average on OECD's "Indicators of Employment Protection" for workers on temporary contracts and above average for workers on permanent contracts (OECD 2013). Moreover, we are not aware of any differences in working conditions for temporary and permanently employed drivers at this particular firm.

What are the practical implications that can be derived from our results? From an organizational point of view, the results seem to suggest keeping employees on temporary contracts as long as possible to maximize the returns from higher effort levels. Yet, labor laws rule out recurring extension of temporary contracts. It may, therefore, appear an even more promising strategy to employ temporary workers only. However, in the absence of promotion opportunities (from temporary to permanent contracts) temporary workers lack the necessary incentives to choose high levels of effort. Dolado and Stucchi (2008) find that firms with a high contract conversion rate (by promoting workers from temporary to permanent contracts) display higher levels of labor productivity. Thus, it seems a reasonable strategy to openly communicate the screening nature of temporary contracts since then temporary workers will be motivated to choose high levels of effort. Moreover, offering properly designed monetary incentives during the weeks and months after contract conversion may help to deter negative behavioral responses by workers. Thus, by offering employees “new” incentives, firms can compensate for the loss of incentives after having signed formerly temporary workers to permanent contracts.

6 SUMMARY AND FUTURE OUTLOOK

This research provides a comprehensive analysis on how economic and social influences affect employee behavior and, hence, worker productivity. Primarily, the aim of the present work is to contribute to a better understanding of the influence that environmental variables – in terms of incentives and teamwork – exert on workers' effort choice decisions in order to identify ways and instruments that may enhance worker productivity. Along these research interests, the four separate analyses of the work at hand can be divided into two parts. The first part (chapters two and three) centers on potential social determinants within work groups (peer effects and worker heterogeneity) that may affect workers' effort choices. Subsequently, the second part (chapter three to five) analyzes the link between potential economic determinants (incentives) and workers' effort choice. Implications to enhance worker productivity are discussed in detail for both parts.

Throughout this thesis, productivity is assessed by means of the three following categories of performance measures:

- overall team performance based on subjective or objective performance measures (chapter 2)
- employee absenteeism (chapter 3)
- objective computerized performance evaluations (chapters 4 and 5)

The respective covariates of teamwork / work groups used in chapters two and three can be categorized as follows:

- diversity in the composition of the group
  - task-related variables (tenure, function, educational background, educational level)
  - non-task related variables (age, gender, culture)
- team size
- consistency of group composition (turnover)
- worker characteristics (share of temporary workers, share of workers who suffer from health impairments)

The respective covariates of incentives applied in chapters three to five are:

- incentives in the economic environment
  - national regulations (sickness benefits, employment protection laws)

    - economic circumstance (prosperity level, (un)employment rate)
- incentives in the employment relationship
    - performance incentives (bonus pay depending on work performance)
    - contract status (temporary vs. permanent employment contracts)

Relevant insights on social and economic influences that determine worker productivity are provided throughout all four studies of the thesis at hand. While the first study (chapter two) is built on existing empirical research in applying state-of-the-art meta-analytic procedures, the subsequent studies (chapters three to five) are drawing on rich – hitherto unavailable – insider econometric data from inside a global automobile manufacturer. The research interest of each chapter has been broken down to simplified one-sentence research questions (as stated in chapter one):

Chapter 2: Is employee behavior and, thus, work group performance affected by work group composition in terms of homogeneity/heterogeneity?

Chapter 3: What are potential social and economic determinants of employee absence behavior at the team-level?

Chapter 4: Do employees respond to unfavorable changes in extrinsic incentive design by adapting their behavior in terms of effort choice and productivity?

Chapter 5: Do employees adjust their effort choice and productivity when being "promoted" from temporary to permanent employment?

The first two questions address peer effects of work group settings on employee behavior. Findings presented in chapter two confirm work group diversity to affect team performance. In support of similarity-attraction and social categorization theories, a negative effect of less task-related age and gender heterogeneity on overall team performance is identified. In line with economic theory, results for highly task-related educational background diversity suggest a positive relation to work group productivity. Team size and team type moderate the diversity-performance relationship. Implications that seek to increase team productivity have to be distinguished along the task of the team. Two major lessons learned can be stated in order to provide practitioners with relevant information on how to manage work group diversity. First, group heterogeneity should be low in teams that work on standardized or routine tasks, e.g. blue-collar work teams. In contrast to this, group heterogeneity – even on less task-related attributes – is valuable for teams whose tasks require creative thinking, problem solving and decision making, e.g. TMT and R&D teams. Second, large teams regularly outperform smaller teams on creative tasks and decision making. Hence, groups performing these tasks benefit from additional members.

In line with general findings on shirking in groups, results of chapter three reveal productivity to decrease with team size. According to experts from inside the company, this is attributable to less extensive monitoring and absence management activities of line managers' as team size increases. Similarly, inconsistencies in group composition as indicated by high turnover result in lower productivity. This can be explained by the lack of social ties and team spirit which negatively influences employees' motivation and willingness to choose high effort levels. Referring to co-worker characteristics, the share of temporary workers in a team is negatively related to effort choice and, thus, offers support for the existence of shirking. The share of workers with any kind of health impairment is not found to influence effort choice decisions and productivity. Three main implications for practitioners can be derived from these results. First, managers need to invest time in developing close social ties with their subordinates. Although costly at first, productivity gains will arise in the long run since absence management is facilitated. Companies should support managers by creating appropriate conditions when it comes to organizational culture. Second, in order to avoid productivity losses that may arise from shirking behavior of permanently employed workers in the presence of temporary workers, firms would be well advised to keep the share of temporary agents at a moderate level. Third, employee turnover is found to increase group absence and, hence, should be kept on a low level.

With regard to economic determinants of employee behavior as addressed in the second research question, behavioral effects induced by incentives are analyzed in chapters three to five. Incentives may arise either from the nature of the employment relation (e.g. performance pay, contract status) or from the given economic environment (e.g. unemployment situation, sickness benefits). The second empirical study of the present thesis (chapter three) provides mixed findings on incentives resulting from the economic environment. In support of personnel economic theory, findings suggest that incentives to shirk are higher with strict employee rights as well as with an increase in national prosperity level. Both effects are attributable to a feeling of job security that incentivizes employees to take the risk of being caught shirking since they value consequences to be negligible. Despite the fact that the relation between national sickness benefits and employee absence is widely recognized (e.g. Frick, Malo 2008), the present work fails to find support for this notion.

This can be explained by the fact that the company under observation levels out differences in national sickness legislation to a certain degree. Any conclusion on this issue would, therefore, be biased. Additionally, findings on the influence of national (un)employment figures on absence are inconclusive. However, it seems difficult to deduce implications for businesses since the economic incentives discussed are mainly policy-driven. Still, it seems to be advantageous to openly communicate the economic situation of the company in order to increase employees' awareness of potential risks. This may discourage workers from shirking behavior.

Questions three and four are closely related since both of them focus on the productivity effects of economic incentives in the employment relationship. Findings of chapter four are counterintuitive to personnel economic conceptions of incentives as positively affecting productivity. Instead, the results reveal that employees display significantly less effort when receiving individual performance bonus pay. These findings support arguments stated in the crowding-out theory that suggests extrinsic incentives to buy off employees' intrinsic motivation. In the case at hand, intrinsic and social motives such as environmental concerns, peer competition as well as concerns for status and reputation serve as adequate explanations. The most important lesson learned from this finding might be the fact that monetary incentives may not serve as a panacea to align the interests of principals and agents under all circumstances. Instead, it is highly recommended to consider the internal and external fit of incentive schemes and general HRM instruments that are in operation.

In contrast to this, findings of chapter five again appear in line with the personnel economics incentive literature since employees on a temporary contract exert more effort and, thus, are more productive than their colleagues on permanent contracts. Yet, this productivity surplus disappears after employees are promoted to permanent contracts. This behavior is attributable to the incentivizing nature of permanent employment contracts that usually come along with higher pay, better working conditions and stricter employment protection compared to temporary contracts. After having been promoted, drivers do no longer yearn for permanent employment and readjust their effort choice decisions. Managers should, therefore, consider properly designed incentives during the phase of contract conversion. Moreover, findings suggest that a company may use temporary contracts as a screening instrument since temporary agents will be incentivized to be more productive given the prospect of being offered a permanent contract.

Given these findings, four main economic insights can be summarized (in line with Ichniowski and Shaw 2013). First, individual worker productivity is (at least partially) determined by peers. Since individuals in teams may be complements concerning their knowledge and skills (e.g. educational background), team performance is usually more than the sum of the individuals' productivity if a team-based production function is assumed. Second, in accordance with previous theoretical and empirical evidence, the work at hand suggests that management practices such as incentive schemes, temporary contracts and teamwork influence workers' effort choice decisions at the intensive margin. Thus, the introduction of any new HRM instrument may be assumed to affect individual worker productivity. Third, in line with previous research, empirical evidence presented throughout this work emphasizes managerial practices and human resource instruments to have a large effect upon worker productivity and, thus, companies' baseline results. Although not all studies of this thesis allow for clear quantitative assessments of productivity shifts, the overall results suggest that employees' effort choice decisions in response to specific human resource practices have a substantial impact on the profitability of the company. Fourth, in line with the general approach of personnel economics, the results presented – in large parts – confirm human beings to behave as rational utility maximizers who aim at increasing their individual income. However, there is also evidence indicating that employees might behave far less incentive-oriented in decision making as presumed. Instead of responding to incentive payments by increasing effort as expected by economic theory, workers are found to reduce individual effort in response to extrinsic incentives. This reduces worker productivity and imposes additional costs on the employer that may be interpreted as hidden costs of incentives. As a result, managers need to carefully evaluate the ideal combination of management practices and HRM instruments that best suits the current situation of the company, the industry, the employees, and the set of HRM instruments in practice. Often management practices are complements of one another and, therefore, might not develop their full potential when only used isolated (e.g. Ichniowski, Shaw, Prennushi 1997, Milgrom, Roberts 1990).

Due to its broad scope with regard to research field, data and method the present work contributes to personnel economics for several reasons. First, the major part of this thesis is built on hitherto unavailable insider data gathered at one international company and, thus, extends the emerging strand of insider econometric literature. Moreover, parts of this work are based on highly innovative GPS-based fleet management data that – to the best of the author’s knowledge – has not yet been used in personnel economics research. Since availability and quality of this objective real-time data is overwhelming, its study may offer broad advantages in analyzing employee behavior. This thesis, therefore, makes a relevant claim to include this innovative and promising type of data into personnel economics. Second, the use of objective performance measures as indicators of worker productivity depicts a distinct advantage over most existing research that is often based merely on subjective (self-) rated performance. This is particularly true for absenteeism research since register absence data is often unavailable to researchers. As a consequence, most existing findings on absence are based on self-reported absence figures that are known to be biased due to underestimation with actual absence being twice as high (Johns 1994). Hence, the objective absence data used in chapter three depicts a clear advantage. Similarly, analyses in chapters four and five benefit from computerized performance evaluations that are collected automatically and objectively rate performance based on predefined algorithms. In general, the use of high quality objective performance measures constitutes a major advantage of the dissertation at hand. It addresses weaknesses which can be found in parts of the existing productivity research that often rely on subjective performance measures. Based on objective performance data, the presented findings complement existing evidence and, thus, contribute to the progress of the discussion in personnel economics. Third, despite the fact that incentives are often framed to be the core of personnel economics (e.g. Lazear 2000b) and as such are probably one of the most intensively studied fields in HRM research, very little is known about responses of employees to the abolition of incentives. To the best of the author’s knowledge, the abolition of incentives has not been studied that often – Freeman and Kleiner (2005) representing a rare exception. By studying employees who experience the abolition of an existing incentive scheme, this work may contribute to a better understanding of the general functioning of incentives. Eventually, this thesis contributes to absence research in three ways. First, as mentioned above the use of register data in absence research depicts a clear advantage over most of the existing research that lacks access to objective absence figures from inside a company. Second, the absence data used in this study is advantageous in that it was compiled in different international plants of the same company. Hence, the data is coherently recorded using identical corporate standards at all plants. This allows for international comparisons without facing the common drawbacks of varying reporting methods usually associated with international absence data (European Foundation for the Improvement of Living and Working Conditions 2010). Third, absence here is primarily evaluated at the group level. Despite the fact that absence is a social concept and absence behavior is to a large extent influenced by worker's peers, personnel economic knowledge on absenteeism is to a great part based on individual-level data (e.g. Rentsch, Steel 2003). Chapter three addresses this gap in absence research. In total, the findings of this thesis contribute to the ongoing discussion of the social and economic determinants in personnel economics by offering important insights on teamwork and incentives based on unique data.

Despite its valuable contribution to personnel economics the work at hand faces some limitations originating from both econometric methods and data. In general, meta-analyses face two particular challenges (e.g. Egger, Smith, Sterne 2001): the publication bias as well as the garbage-in-garbage-out issue. First, studies with positive and significant findings are more likely to be published in peer-reviewed journals. As a result, meta-analyses that focus solely on published articles only cover significant findings and leave out controversial or inconclusive results, leading to the emergence of a publication bias. Second, the quality of any meta-analysis is dependent on the quality of the studies reviewed. To account for this pitfall – referred to as garbage-in-garbage-out effect – only studies published in high-quality peer-reviewed journals are included in the meta-analysis presented in chapter two. Albeit the focus on peer-reviewed articles might provoke a publication bias, this procedure seems to be particularly crucial to identify the most important contributions to the extensive literature on team diversity. Moreover, any publication bias is mitigated by the fact that the findings included in the meta-analysis in chapter two show positive as well as negative correlations.

A common issue of insider studies is centered on its single-firm data sources. Although this allows for high-quality research at the micro-perspective, the generalizability of results is limited (Ichniowski, Shaw 2013). Moreover, this research is likely to suffer from selection bias and, therefore, it also suffers from endogeneity in the choice of workers and managers as it focuses exclusively on a specific industry. Results based solely on data and information originating from the automotive industry may not be presumed to be universally valid and, therefore, should not serve as a blueprint for other industries without further investigation. A distinctive limitation to all non-experimental data – and thus to insider econometrics as well – is that the choice of a treatment by means of a particular management practice is not random but instead is the result of a maximization decision taken by the company.

In conclusion, the evidence and implications discussed in this thesis provide important contributions to the strands of insider econometrics and personnel economics, in particular to the fields of incentive design and work organization in teams. With regard to the methods applied and topics studied, the work at hand concludes by pointing out promising suggestions for future research. First, although having a long tradition in other scientific fields such as medicine, the meta-analytic approach is only slowly gaining credence in economic research. However, results of this thesis demonstrate that meta-analyses are a useful means to overcome limitations of small sample sizes and gain an overview on topics with extensive, but so far inconclusive empirical evidence. In particular, topics that can be assessed as being “over-researched” may benefit more from meta-analyzed conclusions than from additional empirical findings. Second, modern times are changing the circumstances of firms and employees alike as complexity of cooperation and processes increases. Therefore, economic models need to be adapted and refined by incorporating information and data from within companies. On this account, insider studies are a key to improving the quality of both theoretical models and implications for practitioners derived from research. Thus, further contributions to insider econometrics are highly appreciated. Third, in order to gain a comprehensive understanding of an economic phenomenon, it is crucial to analyze it in all its facets. For instance, in response to the well-studied field of incentive introduction, this work analyses the behavioral consequences of its abolition. As a consequence, the presented findings may contribute to an overall better understanding of incentives. Therefore, more studies that depart from familiar paths and shed light on hitherto unstudied aspects of well-known subjects are needed. Fourth, as mentioned earlier, the use of data from GPS-based fleet management systems has proven to contribute significant findings to personnel economics. Thus, scholars should keep their eyes open to new and innovative data sources that so far have not been taken into account for scientific research. The digital age can be assumed to offer further “treasure chests” in terms of computerized data. Eventually, this dissertation emphasizes the advantageousness of the cooperation between academia and businesses in the context of insider econometrics. The author not only acknowledges the great opportunity to work with unique and otherwise unavailable data from inside a company but also highly appreciates the chance to discuss findings with internal experts in order to refine interpretations and derive more comprehensive implications. The author of this thesis, therefore, invites researchers as well as practitioners to join their forces in the search for valuable insights that benefit both scientific progress and companies’ bottom-lines.

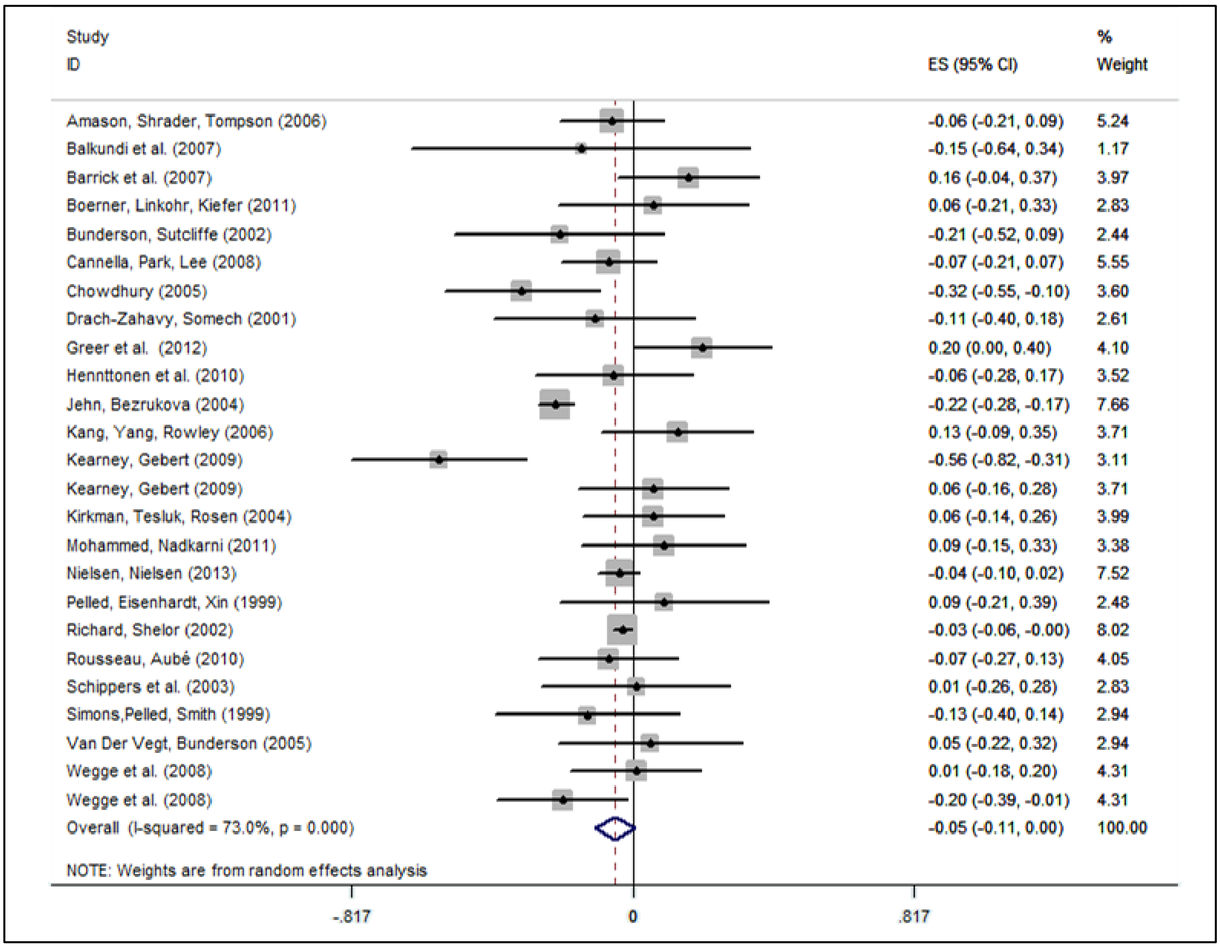
APPENDIX

Table A.1: Heterogeneity Measures Listed by Frequency of Use

Diversity measure	Blau	CV	Teachman	s.d.	Other*
Age	4	14	2	5	3
Gender	14	1	5	1	8
Culture	13	1	3	-	2
Function	14	-	5	-	3
Tenure	2	16	-	4	1
Education					
Background	7	-	-	3	3
Level	4	3	2	-	1

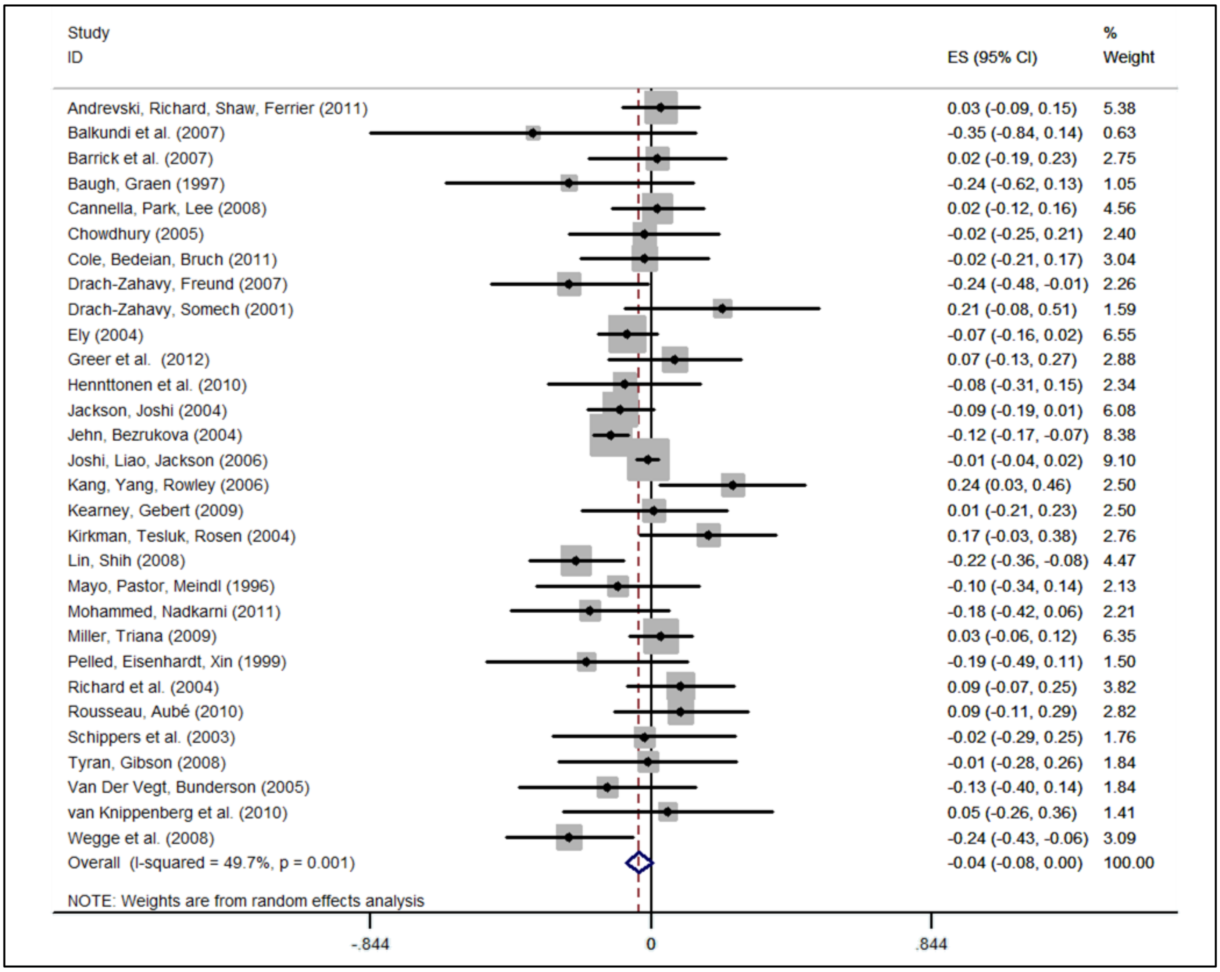
* Other heterogeneity measures include Herfindal index, Gini index, mean, percentage share and authors' own modifications.

Figure A.1 : Forest Plot – Age Diversity and Overall Performance



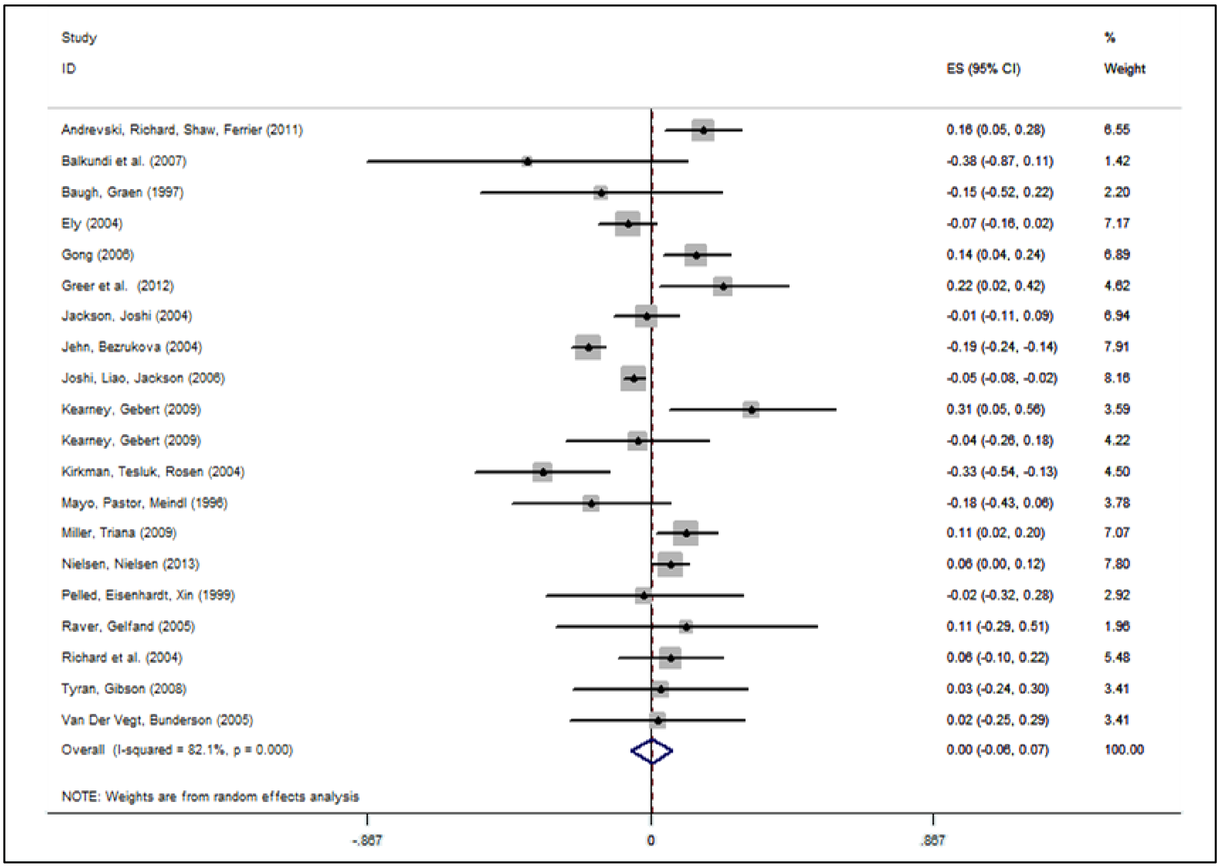
Source: Own calculations.

Figure A.2: Forest Plot – Gender Diversity and Overall Performance



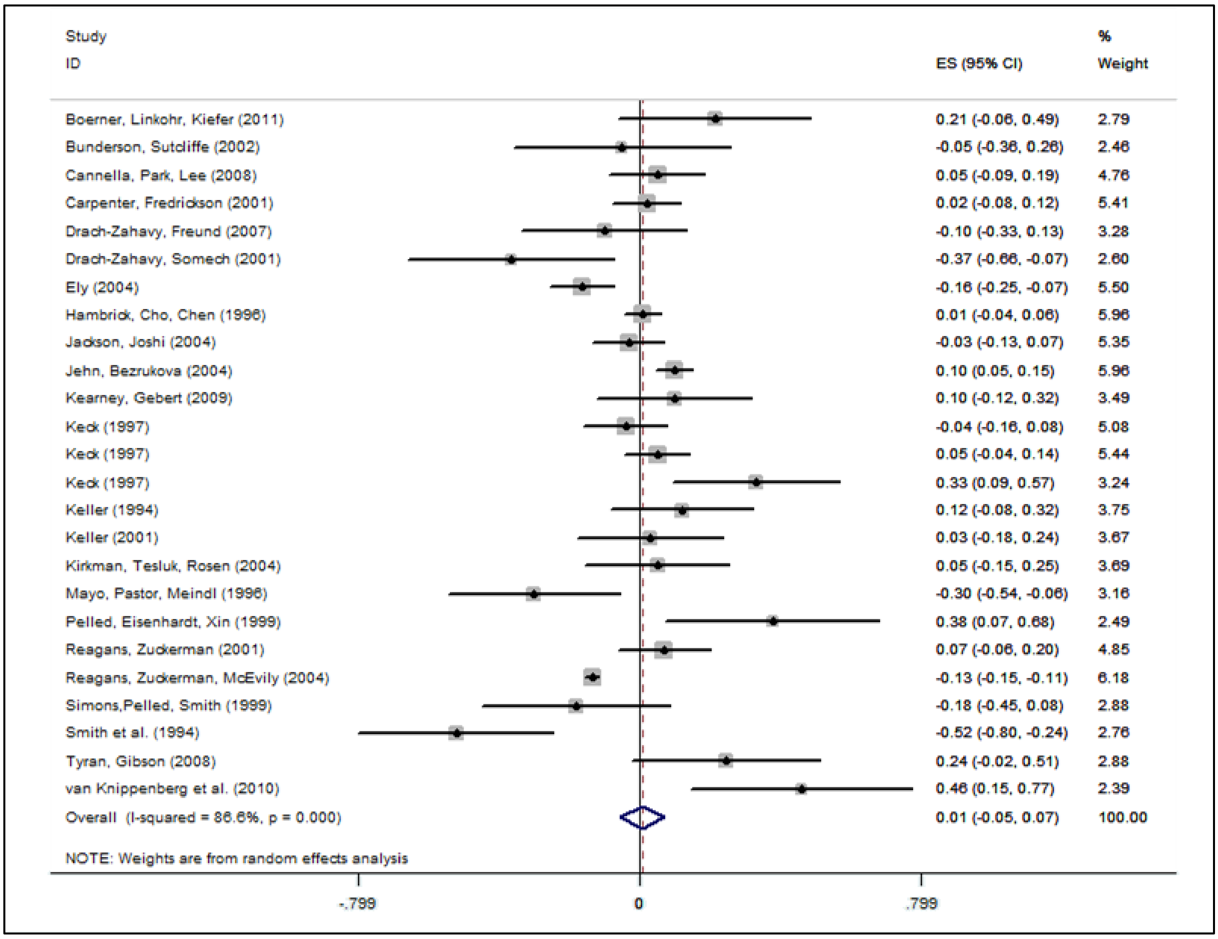
Source: Own calculations.

Figure A.3: Forest Plot – Culture Diversity and Overall Performance



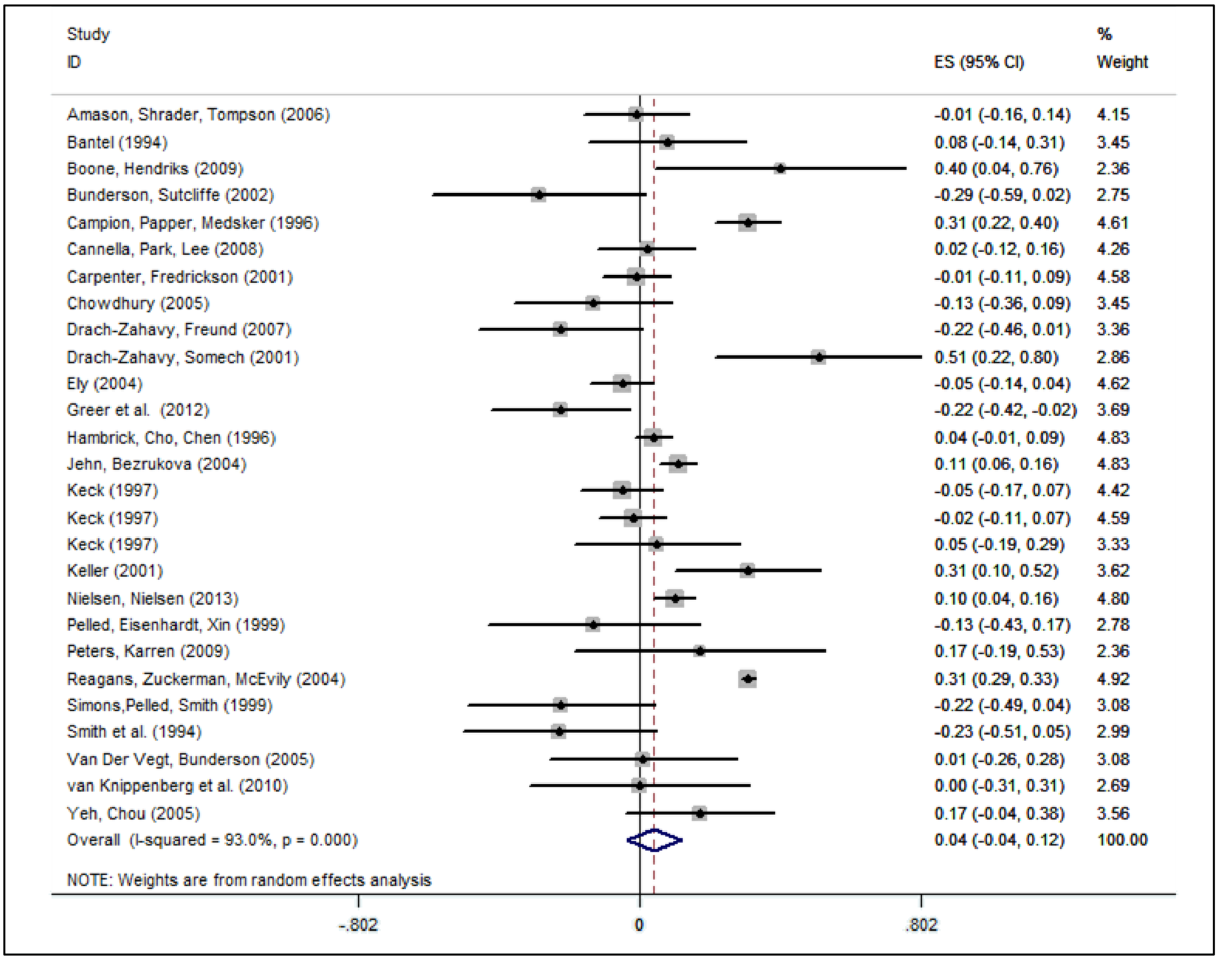
Source: Own calculations.

Figure A.4: Forest Plot – Tenure Diversity and Overall Performance



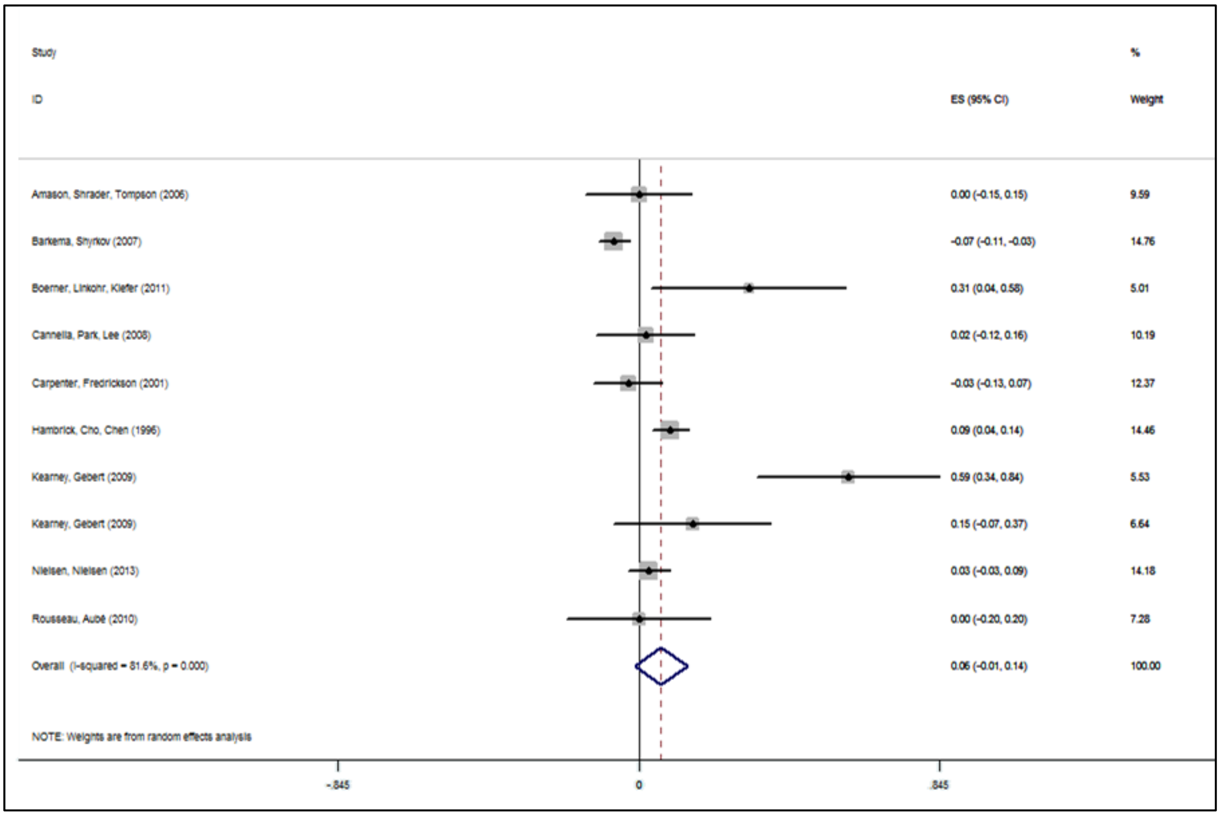
Source: Own calculations.

Figure A.5: Forest Plot – Functional Background Diversity and Overall Performance



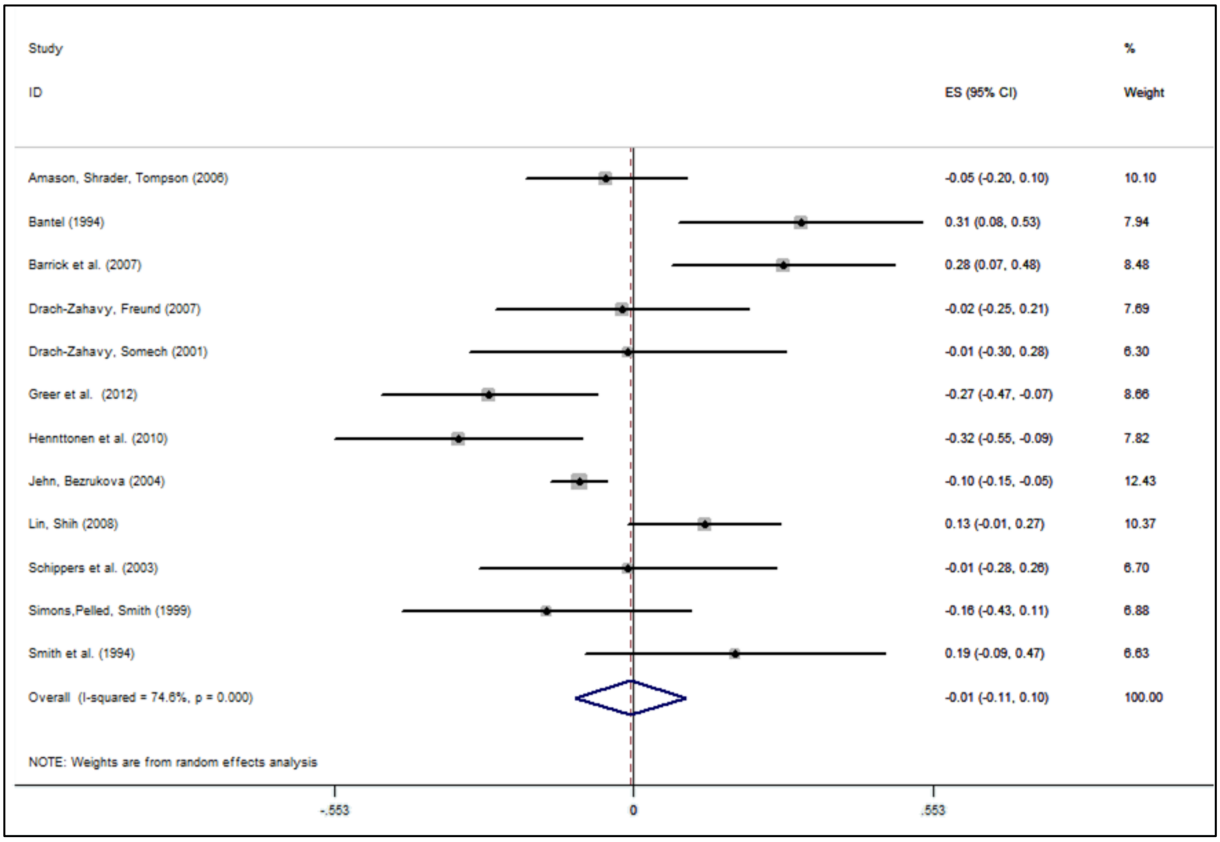
Source: Own calculations.

Figure A.6: Forest Plot – Educational Background Diversity and Overall Performance



Source: Own calculations.

Figure A.7: Forest Plot – Education Level Diversity and Overall Performance



Source: Own calculations.

Table A.2: Diversity-Performance Relation by Performance Measure

Performance Measure	k	r_z	95% CI Lower	95% CI Upper	Q	df	I^2
Age							
Financial	7	-0.029	-0.069	0.011	7.66	6	21.70%
Subjective	12	-0.046	-0.165	0.073	45.3***	11	75.70%
External	10	-0.074	-0.2	0.051	34.31***	9	73.80%
Team	3	0.052	-0.082	0.186	2.02	2	0.80%
Effectiveness	4	-0.073	-0.264	0.118	8.81**	3	66.00%
Gender							
Financial	8	-0.01	-0.042	0.022	7.72	7	9.30%
Subjective	13	-0.064*	-0.13	0.003	17.25	12	30.40%
External	10	-0.078*	-0.162	0.006	15.15*	9	40.60%
Team	4	-0.022	-0.132	0.087	0.41	3	0.00%
Effectiveness	8	-0.073	-0.201	0.054	17.15**	7	59.20%
Culture							
Financial	9	0.045	-0.019	0.109	33.16***	8	75.90%
Subjective	7	-0.085	-0.236	0.066	21.63***	6	72.30%
External	7	-0.085	-0.236	0.066	21.63***		72.30%
Team	---	---	---	---	---	---	---
Effectiveness	4	-0.006	-0.184	0.172	7.44*	3	59.70%
Function							
Financial	13	-0.014	-0.075	0.047	29.23***	12	58.90%
Subjective	5	0.052	-0.11	0.215	27.51***	4	85.50%
External	5	0.052	-0.11	0.215	27.51***	4	85.50%
Team	---	---	---	---	---	---	---
Effectiveness	7	0.086	-0.075	0.246	120.32***	6	95.00%
Tenure							
Financial	10	-0.013	-0.102	0.077	37.27***	9	75.80%
Subjective	6	0.095**	0.021	0.169	6.29	5	20.50%
External	6	0.095**	0.021	0.169	6.29	5	20.50%
Team	---	---	---	---	---	---	---
Effectiveness	8	-0.001	-0.107	0.106	57.28***	7	87.80%
Education							
Background							
Financial	6	0	-0.064	0.063	13.49**	5	62.90%
Subjective	3	0.239	-0.09	0.569	12.97***	2	84.60%
External	3	0.239	-0.09	0.569	12.97***	2	84.60%
Team	---	---	---	---	---	---	---
Effectiveness	1	0.09***	0.038	0.142	0***	0	0.00%

(*continues next page*)

Level							
Financial	4	0.035	-0.206	0.276	16.33***	3	81.60%
Subjective	5	-0.038	-0.204	0.129	16.99***	4	76.50%
External	3	-0.132*	-0.267	0.002	4.03	2	50.30%
Team	3	-0.017	-0.383	0.35	14.63***	2	86.30%
Effectiveness	2	0.008	-0.275	0.29	3.62*	1	72.30%

Note. k = total number of correlation coefficients meta-analyzed; *N* = total number of teams across the correlations; r_z = corrected population correlation (sample-size weighted based ES on Fisher's z transformed correlation coefficients with significance test for ES=0; r_z = standard error of the corrected population correlation; *95% CI* = lower and upper bound of the 95% confidence interval; Q = homogeneity statistics (Cochran's Q); df = degree of freedom; I^2 = percentage of between-study variation due to heterogeneity. Occasions of no or insufficient observations marked with ---.
***p<.01; **p<.05; *p<.1

Table A.3: Fixed-effects Estimations on Absence Rate (IV)+

	luxury sites						volume sites					
	IV-(1)			IV-(2)			IV-(3)			IV-(4)		
VARIABLES	**Plant A (GER)**			**Plant B (UK)**			**Plant C (ESP)**			**Plant D (GER)**		
Mean age (in years)	-1.477	-0.529	-0.197	-0.336	0.402	-0.603	-0.735	0.0102	-0.666	1.342	0.516**	-0.0372
	(7.355)	(0.356)	(7.552)	(3.664)	(0.434)	(3.641)	(1.415)	(0.112)	(1.417)	(2.003)	(0.230)	(2.896)
Mean age² (in years)	0.0124	---	-0.0052	0.0101	---	0.00945	0.0101	---	0.00835	-0.0114	---	0.00668
	(0.101)		(0.104)	(0.038)		(0.0411)	(0.0189)		(0.0188)	(0.025)		(0.0354)
Share of males	-16.63	-12.47	-6.985	8.525	12.53	8.616	-2.601*	-2.487*	-2.493*	-0.0704	1.430	0.00338
(in %/100)	(23.22)	(18.01)	(23.09)	(8.605)	(7.577)	(8.321)	(1.408)	(1.398)	(1.404)	(5.021)	(6.919)	(6.917)
Mean tenure (in years)	3.171	-4.040*	-3.353	-2.738*	-0.135	-1.968	-0.213	0.0111	-0.189	0.515	-0.630	1.630
	(5.626)	(1.879)	(6.395)	(1.332)	(0.221)	(1.179)	(0.400)	(0.129)	(0.402)	(1.144)	(0.432)	(1.653)
Mean tenure² (in years)	-0.521	---	-0.0473	0.0739*	---	0.0561	0.00931	---	0.00740	-0.0464	---	-0.0867
	(0.349)		(0.461)	(0.037)		(0.0343)	(0.0145)		(0.0148)	(0.043)		(0.0611)
Share of temporary	-6.72**	-6.52**	-2.61**	1.034	0.0511	2.329*	n/a	n/a	n/a	-1.397	-1.492	-1.380
workers (in %/100)	(1.911)	(2.371)	(0.903)	(1.691)	(2.297)	(1.200)				(3.374)	(3.392)	(3.4499)
Employment protection	n/v	n/v	---	5.707*	5.040*	---	-1.647	-2.209	---	n/v	n/v	---
legislation (0=weak,…)				(2.515)	(2.218)		(10.81)	(10.83)				
2-Shift system (yes=1)	-7.341	-8.090*	-9.248	2.830**	2.683**	2.052**	2.343*	2.393*	1.763	n/v	n/v	n/v
	(4.008)	(3.834)	(4.723)	(0.974)	(0.850)	(0.744)	(1.293)	(1.286)	(1.185)			
Constant	5.620	26.62	73.24	-5.480	-42.92	25.10	11.83	-3.405	15.57	-16.72	-81.02	-9.533
	(144.1)	(81.92)	(114.7)	(103.6)	(40.28)	(76.52)	(35.02)	(21.28)	(25.37)	(40.03)	(153.7)	(53.10)
Observations	198	198	198	273	273	273	3,464	3,464	3,464	1,414	828	828
R-squared	0.335	0.324	0.290	0.155	0.142	0.137	0.047	0.046	0.041	0.075	0.073	0.075
Number of unit	6	6	6	8	8	8	104	104	104	42	42	42
Environmental Var.	YES	YES	---	YES	YES	---	YES	YES	---	YES	YES	---

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

+ all estimations including month and weather dummies & age and tenure coefficients of variation | only main results displayed | n/a= not available | n/v= no variance

Table A.4: Fixed-effects Estimations on Mean Absence Spell Duration (V)+

	luxury sites						**volume sites**					
	V-(1)			**V-(2)**			**V-(3)**			**V-(4)**		
VARIABLES	**Plant A (GER)**			**Plant B (UK)**			**Plant C (ESP)**			**Plant D (GER)**		
Mean age	0.676	-0.429	-0.607	-6.752	-0.121	-6.115	-2.518	0.0349	-2.447	0.590	0.678**	-2.242
(in years)	(3.048)	(0.398)	(2.784)	(4.676)	(0.508)	(5.018)	(2.291)	(0.138)	(2.290)	(2.684)	(0.302)	(3.375)
Mean age²	-0.0147	---	0.00461	0.0771	---	0.0773	0.0345	---	0.0327	-0.00069		0.0354
(in years)	(0.0430)		(0.0395)	(0.0518)		(0.0579)	(0.0310)		(0.0310)	(0.0329)		(0.041)
Share of males	-20.76*	-24.18*	-15.98	3.859	6.602	7.593	-2.614	-2.350	-2.515	-3.559	-2.779	-3.529
(in %/100)	(10.11)	(11.40)	(14.87)	(8.040)	(8.063)	(7.491)	(1.864)	(1.836)	(1.845)	(6.649)	(8.869)	(8.794)
Mean tenure	-8.708	-2.374	-4.711	-0.543	0.233	-1.325	-0.306	0.129	-0.297	-0.893	-0.847	0.963
(in years)	(5.580)	(1.977)	(5.520)	(1.393)	(0.338)	(1.033)	(0.455)	(0.139)	(0.450)	(1.716)	(0.563)	(2.504)
Influenza activity	-0.416	-0.502	-0.244	-0.174	-0.0660	0.0923	0.734	0.768	0.774*	-0.129	-2.009	-2.031
(0=no activity,…)	(0.578)	(0.619)	(0.554)	(0.596)	(0.479)	(0.577)	(0.623)	(0.624)	(0.450)	(0.524)	(2.824)	(1.539)
Turnover	0.0547	0.0306	-0.0187	0.0227	0.0219	0.0235	-0.0601	-0.0561	-0.0478	-0.604**	-0.59**	-0.59**
(arrivals & exits)	(0.136)	(0.134)	(0.112)	(0.0161)	(0.0131)	(0.0185)	(0.0518)	(0.0519)	(0.0505)	(0.270)	(0.265)	(0.256)
Unit size	-0.0172	0.0320	-0.0169	-0.0256	-0.0804	-0.0143	0.29***	0.28***	0.311***	0.209	0.435*	0.426*
(in persons)	(0.0969)	(0.102)	(0.116)	(0.102)	(0.0744)	(0.113)	(0.0618)	(0.0636)	(0.0600)	(0.146)	(0.245)	(0.234)
2-Shift system	-1.605	-0.963	-2.206	1.142	1.301	1.647*	4.044**	4.141**	3.414**	n/a	n/a	n/a
(yes=1)	(2.981)	(2.663)	(3.303)	(1.111)	(0.850)	(0.725)	(1.758)	(1.768)	(1.594)			
3-Shift system	n/a	n/a	n/a	n/a	n/a	n/a	2.737*	2.841**	2.329*	n/v	n/v	n/v
(yes=1)							(1.420)	(1.431)	(1.361)			
Constant	78.50	65.06	65.48*	113.3	-35.47	137.2	29.70	-20.01	41.40	8.170	74.69	44.16
	(40.34)	(46.93)	(31.65)	(128.6)	(54.27)	(113.7)	(49.65)	(24.45)	(41.02)	(52.84)	(181.1)	(62.32)
Observations	198	198	198	273	273	273	3,464	3,464	3,464	1,353	812	812
R-squared	0.213	0.198	0.178	0.151	0.142	0.124	0.046	0.044	0.044	0.049	0.065	0.067
Number of unit	6	6	6	8	8	8	104	104	104	39	39	39
Environmental Var.	YES	YES	---	YES	YES	---	YES	YES	---	YES	YES	---

Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

+ all estimations including month and weather dummies & age and tenure coefficients of variation | only main results displayed | n/a= not available | n/v= no variance

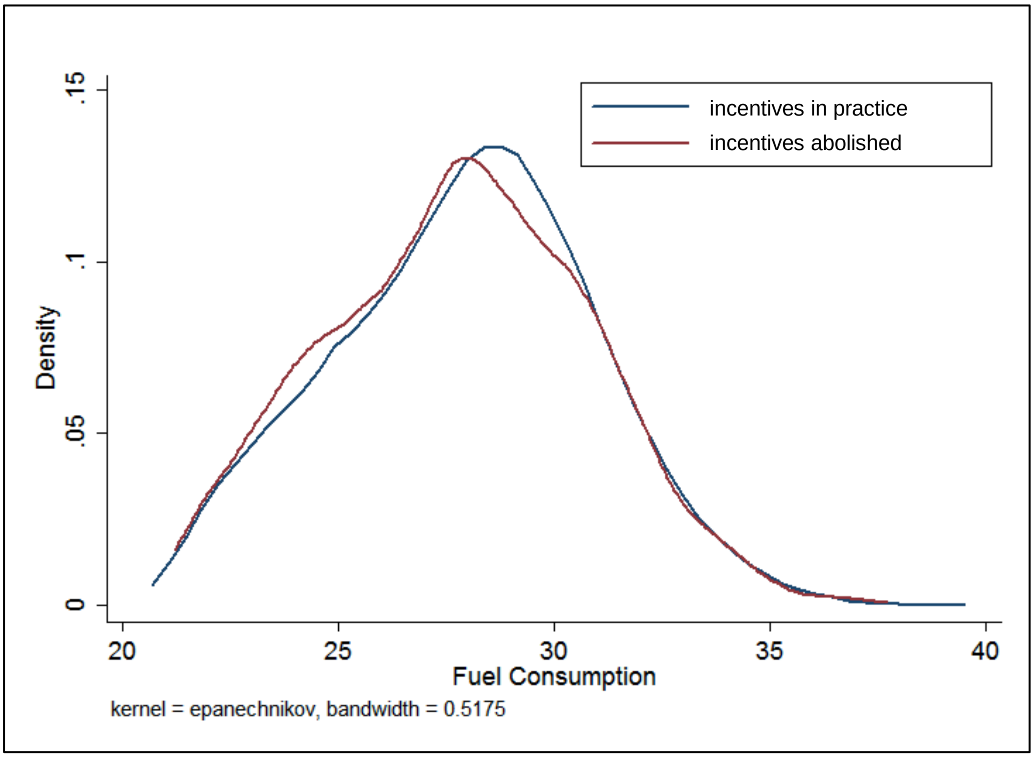
Table A.5: Fixed-effects Estimations on Absence Frequency (VI)+

	luxury sites						volume sites					
	VI-(1)			VI-(2)			VI-(3)			VI-(4)		
VARIABLES	**Plant A (GER)**			**Plant B (UK)**			**Plant C (ESP)**			**Plant D (GER)**		
Mean age	-0.181	0.00743	-0.140	0.117	0.00522	0.129	-0.0132	0.00056	-0.0135	0.0198	---	0.0466
(in years)	(0.174)	(0.0098)	(0.173)	(0.124)	(0.0091)	(0.114)	(0.0235)	(0.0018)	(0.0235)	(0.0466)	(0.004)	(0.0616)
Share of males	1.765*	1.962*	1.832*	-0.0306	0.00402	0.0162	-0.056**	-0.056**	-0.05**	0.0345	0.0455	-0.0006
(in %/100)	(0.805)	(0.776)	(0.775)	(0.332)	(0.343)	(0.295)	(0.026)	(0.0254)	(0.026)	(0.0807)	(0.120)	(0.115)
Mean tenure	0.463**	-0.0588	0.304	-0.0421	-0.0002	-0.0233	0.0021	-0.0004	0.00266	0.0481	-0.002	0.0513
(in years)	(0.138)	(0.0674)	(0.178)	(0.0516)	(0.0053)	(0.0503)	(0.0057)	(0.0022)	(0.0057)	(0.0382)	(0.008)	(0.0412)
Mean tenure²	-0.038**	---	-0.026*	0.00120	---	0.0007	-0.00010	---	-0.0001	-0.0018	---	-0.002
(in years)	(0.0106)		(0.013)	(0.00153)		(0.0015)	(0.0002)		(0.0002)	(0.0014)		(0.0015)
Turnover	-0.0038	-0.0017	0.00032	0.000206	0.00023	0.00018	0.0006	0.00059	0.00085	0.023**	0.015*	0.0169*
(arrivals & exits)	(0.0051)	(0.0059)	(0.005)	(0.00026)	(0.0003)	(0.0003)	(0.0665)	(0.0652)	(0.0663)		(0.06)	(0.0450)
Share of temporary	0.302**	0.319*	0.296*	0.00786	-0.0179	0.0420	n/a	n/a	n/a	-0.012	-0.0204	0.0037
workers (in %/100)	(0.0965)	(0.132)	(0.126)	(0.0642)	(0.0855)	(0.0588)				(0.064)	(0.068)	(0.0606)
Unit size	0.00028	-0.0031	-0.0028	-0.00317*	-0.002*	-0.0030	0.00034	0.00036	0.00069	-0.0033	-0.0044	-0.0029
(in persons)	(0.0032)	(0.0039)	(0.004)	(0.00165)	(0.001)	(0.0017)	(0.0007)	(0.0007)	(0.0007)	(0.0036)	(0.004)	(0.0036)
Employment prot,	n/v	n/v	---	0.0939*	0.0722	---	-0.230	-0.231	---	n/v	n/v	---
legislation (0=weak)				(0.0403)	(0.0457)		(0.205)	(0.206)				
2-Shift system	-0.0677	-0.114	-0.160*	0.0742*	0.0679	0.0528	0.0316*	0.0310*	0.035**	n/a	n/a	n/a
(yes=1)	(0.0830)	(0.120)	(0.075)	(0.0351)	(0.0359)	(0.0311)	(0.0178)	(0.0176)	(0.0171)			
Constant	1.885	1.207	0.176	-2.522	-0.467	-2.420	0.241	0.0107	0.280	-0.365	-2.541	-1.052
	(4.337)	(3.444)	(2.755)	(3.026)	(0.895)	(2.336)	(0.573)	(0.364)	(0.427)	(1.067)	(5.782)	(1.144)
Observations	198	198	198	273	273	273	3,460	3,460	3,460	1,350	809	809
R-squared	0.388	0.347	0.357	0.230	0.221	0.206	0.076	0.075	0.069	0.086	0.072	0.073
Number of unit	6	6	6	8	8	8	104	104	104	39	39	39
Environmental Var.	YES	YES	---	YES	YES	---	YES	YES	---	YES	YES	---

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

+ all estimations including month and weather dummies & age and tenure coefficients of variation | only main results displayed | n/a= not available | n/v= no variance

Figure A.8: Fuel Consumption with / without Incentives



Source: Own calculations.

Table A.6: Descriptive Statistics with and without Incentives in Practice

	Incentives in Practice[+]				***Incentives Abolished***			
Variable	***mean***	***s.d.***	***between***	***within***	***mean***	***s.d.***	***between***	***within***
Fuel Consumption (in l/100 km)	27.93	3.01	.32	2.99	27.77	3.04	.41	3.02
Driving Parameters								
Average Speed (in km/h)	75.17	3.50	.33	3.48	75.62	3.07	.48	3.03
Idling (in % of engine running time)	.03	.02	.002	.02	.02	.02	.002	.02
Coasting (in % of engine running time)	.14	.05	.01	.05	.15	.05	.01	.05
Speeding (in % of engine running time)	.08	.12	.1	.12	.07	.11	.02	.10
Brake Applications (# per 100km)	12.85	8.28	.88	8.24	11.68	6.91	1.37	6.81
Harsh Brake Applications (# per 100km)	.09	.26	.02	.26	.14	.25	.04	.25
Harsh Accelerations (# per 100km)	.06	.24	.02	.24	.08	.29	.03	.29
Maximum Vehicle Speed (in km/h)	96.10	5.41	.56	5.38	96.36	5.67	1.07	5.59
Driving Scores								
Anticipation (Score in %/100)	.81	.15	.2	.15	.84	.15	.05	.14
Choice Gear (Score in %/100)	.98	.12	.02	.12	.97	.15	.03	.15
Use Brakes (Score in %/100)	.77	.19	.03	.19	.79	.19	.05	.19
Hill Drive (Score in %/100)	.63	.25	.02	.25	.65	.26	.06	.25
Trip Evaluation								
Overall (1=green, 2=yellow, 3=red)	2.04	.99	.37	.92	1.73	.90	.29	.85
Number of drivers: 37								

[+]Excluding the first three months of the observation period due to secretly introduced incentive.

Table A.7: Two-Way Anova of Incentives and Driver on Fuel Consumption

```
                     Number of obs =     6326     R-squared     =  0.0119
                     Root MSE      = 3.02317      Adj R-squared =  0.0004

          Source   Partial SS    df       MS            F     Prob > F

           Model   690.260327    73  9.45562092       1.03     0.3977

      incentives   39.4089719     1  39.4089719       4.31     0.0379
          driver   317.943075    36  8.83175208       0.97     0.5263
incentives#driver  330.581371    36  9.18281587       1.00     0.4610

        Residual   57140.5275  6252  9.13955974

           Total   57830.7878  6325  9.14320756
```

Table A.8: Findings on Fuel Consumption after Installation of Incentives+

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Incentives (yes=1)	1.202	1.168	1.896*	1.846	2.223*
	(.967)	(.977)	(.985)	(1.181)	(1.231)
Tenure (in months)				.0480	.555
				(.612)	(1.019)
Trend (in months after incentive introduction)					-.896
					(1.158)
Constant	27.39***	27.09***	25.82***	25.38***	20.70***
	(.567)	(.575)	(.767)	(5.530)	(9.333)
Vehicle Controls	YES	YES	YES	YES	YES
Date Controls	YES	YES	YES	YES	YES
Time Controls		YES	YES	YES	YES
Weather Controls			YES	YES	YES
Driver FE	YES	YES	YES	YES	YES
Observations	1,199	1,199	1,199	1,199	1,199
R-squared	.082	.084	.093	.093	.093
Number of driver	31	31	31	31	31

Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.1

+ Results of fixed-effects estimations on fuel consumption for short-term effects after the installation of incentives (January 2011 to August 2011, i.e. 3 months without incentives and 5 months with)

Table A.9: Findings on Fuel Consumption after Abolition of Incentive+

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Incentives (yes=1)	.380**	.358**	.338*	.964***	.148	.148
	(.174)	(.174)	(.178)	(.265)	(.388)	(.391)
Age (in years)					-.797***	-.797***
					(.207)	(.209)
Tenure (in months)				.0539**	.0508**	.0508**
				(.0234)	(.0238)	(.0239)
Constant	27.55***	27.21***	26.93***	25.27***	60.75***	63.69***
	(.769)	(.755)	(.806)	(1.091)	(9.770)	(10.54)
Vehicle Controls	YES	YES	YES	YES	YES	YES
Date Controls	YES	YES	YES	YES	YES	YES
Time Controls		YES	YES	YES	YES	YES
Weather Controls			YES	YES	YES	YES
Driver Controls						YES
Driver FE	YES	YES	YES	YES	YES	
Observations	3,263	3,263	3,263	3,263	3,263	3,263
R-squared	.038	.039	.040	.041	.041	
Number of driver	37	37	37	37	37	37

Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.1
+Results of fixed-effects (models 1 to 5) and random-effects (model 6) estimations on fuel consumption for short-term effects after the abolition of the incentives (May 2012 to August 2013, i.e. 8 months with and without bonus each)

Table A.10: Ordered Logit Estimation for Trip Evaluation+

Variables	Ordered Logit Regression
Incentives (yes=1)	1.187***
	(.128)
Age (in years)	.196
	(.226)
Tenure (in months)	.00132
	(.0187)
Vehicle Controls	YES
Date Controls	YES
Time Controls	YES
Weather Controls	YES
Driver Controls	YES
Observations	6,319
Pseudo R-squared	.1136

+Trip Evaluation: “good”=1, “mediocre”=2, “poor”=3
Standard errors in parentheses
*** p<.01, ** p<.05, * p<.1
+(April 2011 to December 2013)

Table A.11: Baseline Results of Fixed Effects Estimation (incl. Bonus Regime)

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Temporary Contract (yes=1)	-.397* (.180)	-.415** (.161)	-.475** (.142)	-.487** (.150)	-.487** (.150)
Bonus Regime (yes=1)	-2.753 (2.608)	-1.772 (2.475)	.436 (1.617)	-1.995 (3.164)	
Tenure (in three-month intervals)				-.074 (.208)	-.074 (.208)
Age (in years)					-.604 (1.550)
Date Dummies	YES	YES	YES	YES	YES
Vehicle Dummies		YES	YES	YES	YES
Time Dummies			YES	YES	YES
Weather Dummies			YES	YES	YES
Driver FE	YES	YES	YES	YES	YES
Constant	28.26*** (2.323)	27.43*** (2.179)	19.04 (18.73)	-4.773 (5.389)	39.11 (64.72)
Number of observations	1,299	1,299	1,299	1,299	1,299
Number of drivers	8	8	8	8	8
R-squared	.090	.116	.128	.128	.128

Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.10

Table A.12: Baseline Results of Fixed Effects Estimation (incl. Trend Variable)

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Temporary Contract (yes=1)	-.675**	-.651***	-.646***	-.657**	-.657**
	(.236)	(.181)	(.181)	(.203)	(.203)
Trend (in month)	-.147	-.137*	-.140*	-.139*	-.139*
	(.081)	(.068)	(.069)	(.070)	(.070)
Trend² (in month)	.00521**	.00452**	.00378*	.00380*	.00380*
	(.00162)	(.00156)	(.00166)	(.00171)	(.00171)
Tenure (in three-month intervals)				-.057	-.057
				(.213)	(.213)
Age (in years)					-.101
					(1.271)
Date Dummies	YES	YES	YES	YES	YES
Vehicle Dummies		YES	YES	YES	YES
Time Dummies			YES	YES	YES
Weather Dummies			YES	YES	YES
Driver Fixed-Effects	YES	YES	YES	YES	YES
Constant	25.74***	25.83***	21.02	19.16	22.65
	(.991)	(1.124)	(12.59)	(13.84)	(54.67)
Number of observations	1,299	1,299	1,299	1,299	1,299
Number of drivers	8	8	8	8	8
R-squared	.092	.117	.130	.130	.130

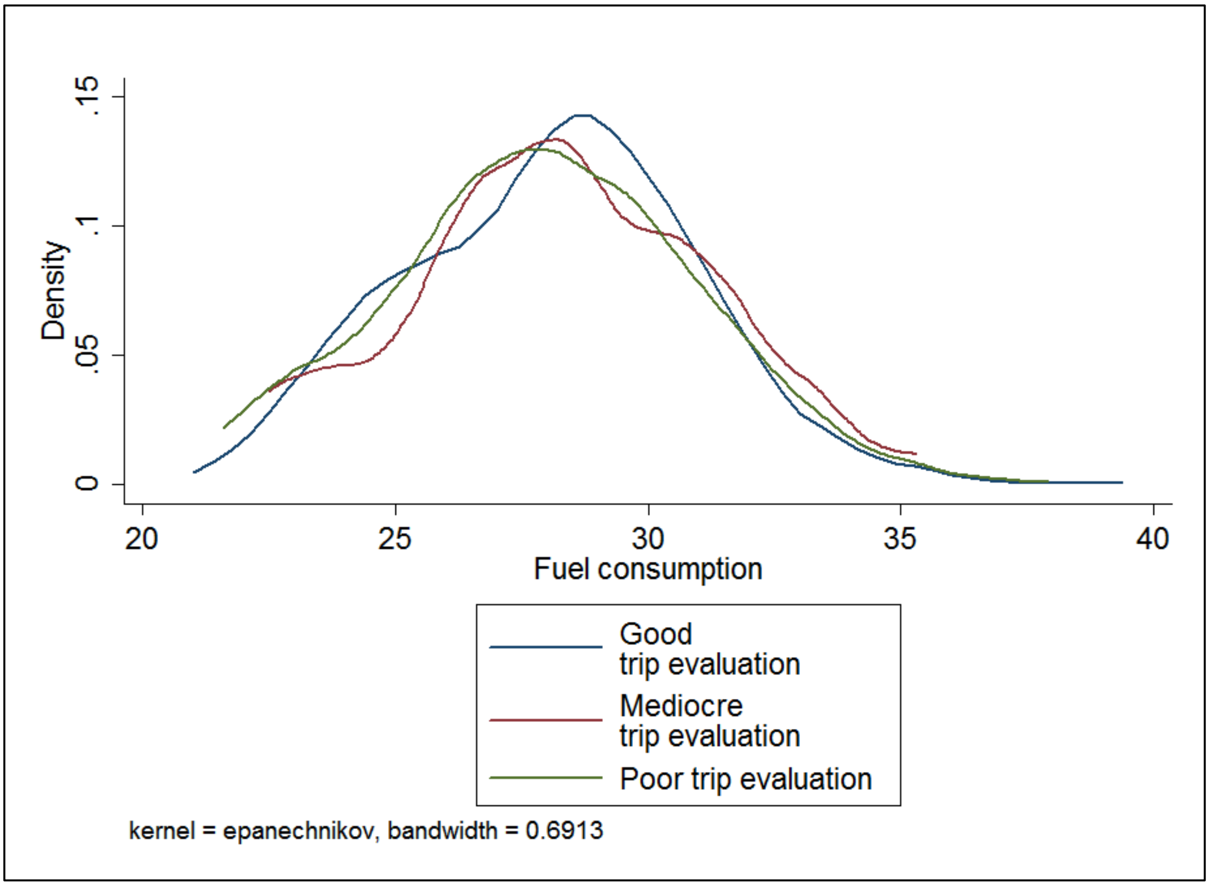
Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.10

Table A.13: Ordered Logit Estimation of Driver Performance[+]

Variables	Ordered Logit
Temporary Contract (yes=1)	-1.4307***
	(.3528)
Age (in years)	.0138
	(.0854)
Tenure (in three-month intervals)	-.1712
	(.3115)
Vehicle Dummies	YES
Date Dummies	YES
Time Dummies	YES
Weather Dummies	YES
Driver Dummies	YES
Number of observations	1,296
Number of drivers	8
Pseudo R-squared	.2955

[+]Driver Performance=Trip Evaluation: “good”=1, “mediocre”=2, “poor”=3
Robust standard errors in parentheses
*** p<.01, ** p<.05, * p<.10

Figure A.9: Kernel Density Plot of Fuel Consumption by Trip Evaluation



Source: Own calculations.

REFERENCES

References marked with asterisks indicate studies that contribute correlations to the meta-analysis presented in chapter two.

Agius, Raymond M.; Lloyd, Margaret H.; Campbell, Susan B.; Hutchinson, Paula A.; Seaton, Anthony; Soutar, Colin A. (1994): Questionnaire for the Identification of Back Pain for Epidemiological Purposes. Occupational and Environmental Medicine, 51(11): 756-760.

Akazawa, Manabu; Sindelar, Jody L.; Paltiel, A. David (2003): Economic Costs of Influenza Related Work Absenteeism. Value in Health, 6(2): 107-115.

Åkerlind, Ingemar; Alexanderson, Kristina; Hensing, Gunnel; Leijon, Margareta; Bjurulf, Per (1996): Sex Differences in Sickness Absence in Relation to Parental Status. Scandinavian Journal of Public Health, 24(1): 27-35.

Åkerstedt, Torbjörn (2003): Shift Work and Disturbed Sleep/Wakefulness. Occupational Medicine, 53(2): 89-94.

Alchian, Armen A.; Demsetz, Harold (1972): Production, Information Costs, and Economic Organization. American Economic Review, 62(5): 777-795.

Alexander, Ralph A.; Scozzaro, Michael J.; Borodkin, Lawrence J. (1989): Statistical and Empirical Examination of the Chi-Square Test for Homogeneity of Correlations in Meta-analysis. Psychological Bulletin, 106(2): 329-331.

Alexanderson, Kristina; Sydsjö, Adam; Hensing, Gunnel; Sydsjö, Gunilla; Carstensen, John (1996): Impact of Pregnancy on Gender Differences in Sickness Absence. Scandinavian Journal of Public Health, 24(3): 169-176.

Ali, Muhammad; Kulik, Carol T.; Metz, Isabel (2011): The Gender Diversity–Performance Relationship in Services and Manufacturing Organizations. International Journal of Human Resource Management, 22(7): 1464-1485.

Allen, Steven G. (1981a): An Empirical Model of Work Attendance. Review of Economics and Statistics, 63(1): 77-87.

Allen, Steven G. (1981b): Compensation, Safety, and Absenteeism: Evidence from the Paper Industry. Industrial and Labor Relations Review, 34(2): 207-218.

Allison, Paul D. (1978): Measures of Inequality. American Sociological Review, 43(6): 865-880.

Amason, Allen C.; Sapienza, Harry J. (1997): The Effects of Top Management Team Size and Interaction Norms on Cognitive and Affective Conflict. Journal of Management, 23(4): 495-516.

*Amason, Allen C.; Shrader, Rodney C.; Tompson, George H. (2006): Newness and Novelty: Related Top Management Team Composition to New Venture Performance. Journal of Business Venturing, 21(1): 125-148.

Amilon, Anna; Wallette, Mårten (2009): Work Absence – A Signalling Factor for Temporary Workers? Labour, 23(1): 171-194.

Amuedo-Dorantes, Catalina (2002): Work Safety in the Context of Temporary Employment: The Spanish Experience. Industrial and Labor Relations Review, 55(2): 262-285.

Ancona, Deborah, G.; Caldwell, David F. (1992): Demography and Design: Predictors of New Product Team Performance. Organization Science, 3(3): 321-341.

*Andrevski, Goce; Richard, Orlando C.; Shaw, Jason D.; Ferrier, Walter J. (2011): Racial Diversity and Firm Performance: The Mediating Role of Competitive Intensity. Journal of Management, 40(3): 820-844.

Arai, Mahmood; Thoursie, Peter S. (2005): Incentives and Selection in Cyclical Absenteeism. Labour Economics, 12(2): 269-280.

Argote, Linda; Gruenfeld, Deborah; Naquin, Charles (2001): Group Learning in Organizations. In: Turner, Marlene E. (ed.) (2001): Groups at Work: Theory and Research. Mahwah (NJ): Lawrence Erlbaum Associates Publishers, pp. 369-411.

Ariely, Dan; Bracha, Anat; Meier, Stephan (2009): Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. American Economic Review, 99(1): 544-555.

Aronsson, Gunnar (1999): Contingent Workers and Health and Safety. Work, Employment & Society, 13(3): 439-459.

Arrunada, Benito; Gonzalez-Diaz, Manuel; Fernandez, Alberto (2004): Determinants of Organizational Form: Transaction Costs and Institutions in the European Trucking Industry. Industrial and Corporate Change, 13(6): 867-882.

Artuç, Erhan; Docquier, Frédéric; Özden, Çaglar; Parsons, Christopher (2015): A Global Assessment of Human Capital Mobility: The Role of Non-OECD Destinations. World Development, 65: 6-26.

Arulampalam, Wiji; Booth, Alison L. (1998): Training and Labour Market Flexibility: Is There a Trade-Off? British Journal of Industrial Relations, 36(4): 521-536.

Arulampalam, Wiji; Bryan, Mark L.; Booth, Alison L. (2004): Training in Europe. Journal of the European Economic Association, 2(2/3): 346-360.

Askildsen, Jan E.; Bratberg, Espen; Nilsen, Øivind A. (2005): Unemployment, Labor Force Composition and Sickness Absence: A Panel Data Study. Health Economics, 14(11): 1087-1101.

Audas, Rick; Goddard, John (2001): Absenteeism, Seasonality, and the Business Cycle. Journal of Economics and Business, 53(4): 405-419.

Azar, Ofer H.; Yosef, Shira; Bar-Eli, Michael (2015): Restaurant Tipping in a Field Experiment: How Do Customers Tip when They Receive Too Much Change? Journal of Economic Psychology, 50: 13-21.

Bacharach, Samuel B.; Bamberger, Peter; Biron, Michal (2010): Alcohol Consumption and Workplace Absenteeism: The Moderating Effect of Social Support. Journal of Applied Psychology, 95(2): 334-348.

Bachler, Christopher J. (1995): Workers Take Leave of Job Stress. Personnel Journal, 74(1): 38-45.

*Baer, Markus; Vadera, Abhijeet K.; Leenders, Roger T.; Oldham, Greg R. (2013): Intergroup Competition As a Double-Edged Sword: How Sex Composition Regulates the Effects of Competition On Group Creativity. Organization Science, 25(3): 892-908.

Baker, George (1992): Incentive Measures and Performance Measurement. Journal of Political Economy, 100(3): 598-614.

Baker, George P.; Hubbard, Thomas N. (2004): Contractibility and Asset Ownership: Onboard Computers and Governance in U.S. Trucking. Quarterly Journal of Economic, 119(4): 1443-1479.

Balafoutas, Loukas; Beck, Adrian; Kerschbamer, Rudolf; Sutter, Matthias (2013): What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods. Review of Economic Studies, 80(3): 876-891.

*Balkundi, Prasad; Kilduff, Martin; Michael, Judd H. (2007): Demographic Antecedents and Performance Consequences of Structural Holes in Work Teams. Journal of Organizational Behavior, 28(2): 241-260.

Bamberger, Peter; Biron, Michal (2007): Group Norms and Excessive Absenteeism: The Role of Peer Referent Others. Organizational Behavior and Human Decision Processes, 103(2): 179-196.

Bandiera, Oriana; Barankay, Iwan; Rasul, Imran (2011): Field Experiments with Firms. Journal of Economic Perspectives, 25(3): 63-82.

Bandiera, Oriana; Barankay, Iwan; Rasul, Imran (2010): Social Incentives in the Workplace. Review of Economic Studies, 77(2): 417-458.

Bandiera, Oriana; Barankay, Iwan; Rasul, Imran (2009): Social Connections and Incentives in the Workplace: Evidence from Personnel Data. Econometrica, 77(4): 1047-1094.

Bandiera, Oriana; Barankay, Iwan; Rasul, Imran (2007): Incentives for Managers and Inequality among Workers: Evidence from a Firm Level Experiment. Quarterly Journal of Economics, 122 (2): 729–773.

Bandiera, Oriana; Barankay, Iwan; Rasul, Imran (2005): Social Preferences and the Response to Incentives: Evidence from Personnel Data. Quarterly Journal of Economics, 120(3): 917-962.

*Bantel, Karen A. (1994): Strategic Planning Openness: The Role of Top Team Demography. Group and Organization Management, 19(4): 406-424.

*Barkema, Harry G.; Shvyrkov, Oleg (2007): Does Top Management Team Diversity Promote or Hamper Foreign Expansion? Strategic Management Journal, 28(7): 663-680.

Barla, Philippe; Bolduc, Denis; Boucher, Nathalie; Watters, Jonathan (2010): Information Technology and Efficiency in Trucking. Canadian Journal of Economics, 43(1): 254-279.

Barmby, Tim A.; Ercolani, Marco G.; Treble, John G. (2002): Sickness Absence: An International Comparison. Economic Journal, 112(480): F315-F331.

Barmby, Tim A.; Orme, Chris D.; Treble, John G. (1995): Worker Absence Histories: A Panel Data Study. Labour Economics, 2(1): 53-65.

Barmby, Tim A.; Sessions, John G.; Treble, John G. (1994): Absenteeism, Efficiency Wages and Shirking. Scandinavian Journal of Economics, 96(4): 561-566.

Baron, James N.; Kreps, David M. (1999): Strategic Human Resources: Frameworks for General Managers. New York (NY): Wiley.

*Barrick, Murray R.; Bradley, Bret H.; Kristof-Brown, Amy L.; Colbert, Amy E. (2007): The Moderating Role of TMT Interdependence: Implications for Real Teams and Working Groups. Academy of Management Journal, 50(3): 544-557.

Barsade, Sigal G.; Ward, Andrew J.; Turner, Jean D.; Sonnenfeld, Jeffrey A. (2000): To Your Heart's Content: A Model of Affective Diversity in Top Management Teams. Administrative Science Quarterly, 45(4): 802-836.

Bartel, Ann; Ichniowski, Casey; Shaw, Kathryn (2004): Using "Insider Econometrics" to Study Productivity. American Economic Review, 94(2): 217-223.

BAuA (2015): Volkswirtschaftliche Kosten durch Arbeitsunfähigkeit 2013. Available at: http://www.baua.de/de/Informationen-fuer-die-Praxis/Statistiken/Arbeitsunfaehigkeit/Kosten.html
Accessed October 13 2014.

*Baugh, S. Gayle; Graen, George B. (1997): Effects of Gender and Racial Composition on Perceptions of Team Performance in Cross-Functional Teams. Group and Organization Management, 22(3): 366-383.

Beal, Daniel J.; Cohen, Robin R.; Burke, Michael J.; McLendon, Christy L. (2003): Cohesion and Performance in Groups: A Meta-Analytic Clarification of the Construct Relationships. Journal of Applied Psychology, 88(6): 989-1004.

Bedeian, Arthur G.; Mossholder, Kevin W. (2000): On the Use of the Coefficient of Variation as A Measure of Diversity. Organizational Research Methods, 3(3): 285-297.

Beemsterboer, Willibrord; Stewart, Roy; Groothoff, Johan; Nijhuis, Frans (2009): A Literature Review on Sick Leave Determinants (1984-2004). International Journal of Occupational Medicine and Environmental Health, 22(2): 169-179.

Bekker, Marrie H.; Rutte, Christel G.; Van Rijswijk, Karen (2009): Sickness Absence: A Gender-Focused Review. Psychology, Health and Medicine, 14(4): 405-418.

Bell, Suzanne T. (2007): Deep-Level Composition Variables as Predictors of Team Performance: A Meta-Analysis. Journal of Applied Psychology, 92(3): 595–615.

Bell, Suzanne T.; Villado, Anton J.; Lukasik, Marc A.; Belau, Larisa; Briggs, Andrea L. (2010): Getting Specific about Demographic Diversity Variable and Team Performance Relationships: A Meta-Analysis. Journal of Management, 37(3): 709-743.

Bénabou, Roland; Tirole, Jean (2003): Intrinsic and Extrinsic Motivation. Review of Economic Studies, 70(3): 489-520.

Bénabou, Roland; Tirole, Jean (2006): Incentives and Prosocial Behavior. American Economic Review, 96(5): 1652-1678.

Benavides, Fernando G.; Benach, Joan (1999): Precarious Employment and Health-Related Outcomes in the European Union. Luxembourg: Office for Official Publications of the European Communities.

Benavides, Fernando G.; Benach, Joan; Diez-Roux, Ana V.; Roman, Carmen (2000): How Do Types of Employment Relate to Health Indicators? Findings from the Second European Survey on Working Conditions. Journal of Epidemiology and Community Health, 54(7): 494-501.

Berton, Fabio; Devincienti, Francesco; Pacelli, Lia (2011): International Journal of Manpower, 32(8): 879-899.

Besley, Timothy; Ghatak, Maitreesh (2005): Competition and Incentives with Motivated Agents. American Economic Review, 95(3): 616-636.

Biemann, Torsten; Kerney, Eric (2010): Size Does Matter: How Varying Group Sizes in a Sample Affect the Most Common Measures of Group Diversity. Organizational Research Methods, 13(3): 582-599.

Bilig, Michael; Tajfel, Henri (1973): Social Categorization and Similarity in Intergroup Behaviour. European Journal of Social Psychology, 3(1): 27-52.

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

Bloom, Nicholas; Eifert, Benn; Mahajan, Aprajit; McKenzie, David; Roberts, John (2013): Does Management Matter? Evidence from India. Quarterly Journal of Economics, 128(1): 1-51.

Bloom, Nicholas; Liang, James; Roberts John; Ying, Zhichun (2015): Does Working from Home Work? Evidence from a Chinese Experiment. Quarterly Journal of Economics, 130(1): 165-218.

Bloom, Nicholas; Van Reenen, John (2011): Human Resource Management and Productivity. In Ashenfelter, Orley; Cards, David (eds.): Handbook of Labor Economics Volume 4b, ed.: Amsterdam, Netherlands: North Holland. 1697-1767.

*Boerner, Sabine; Linkohr, Marius; Kiefer, Sabine (2011): Top Management Team Diversity: Positive in the Short Run, But Negative in the Long Run? Team Performance Management: An International Journal, 17(7/8): 328-353.

Boning, Brent; Ichniowski, Casey; Shaw, Kathryn (2007): Opportunity Counts: Teams and the Effectiveness of Production Incentives. Journal of Labor Economics, 25(4): 613-650.

*Boone, Christophe; Hendriks, Walter (2009): Top Management Team Diversity and Firm Performance: Moderators of Functional Background and Locus of Control Diversity. Management Science, 55(2): 165-180.

Booth, Alison L.; Francesconi, Marco; Frank, Jeff (2002): Temporary Jobs: Stepping Stones Or Dead Ends? Economic Journal, 112(480): F189-FF213.

Booth, Alison L.; Frank, Jeff (1999): Earnings, Productivity, and Performance-Related Pay. Journal of Labor Economics, 17(3): 447-463.

Bowers, Clint, A.; Pharmer, James A.; Salas, Eduardo (2000): When Member Homogeneity is Needed in Work Teams. Small Group Research, 31(3): 305-327.

Boyce, Anthony S.; Ryan, Ann M.; Imus, Anna L.; Morgeson, Frederick P. (2007): “Temporary Worker, Permanent Loser?” A Model of the Stigmatization of Temporary Workers. Journal of Management, 33(5): 5-29.

Bradley, Steve; Green, Colin; Leeves, Gareth (2007): Worker Absence and Shirking: Evidence from Matched Teacher-School Data. Labour Economics, 14(3): 319-334.

Bradley, Steve; Green, Colin; Leeves, Gareth (2012): Employment Protection, Threat and Incentive Effects on Worker Absence. British Journal of Industrial Relations, 52(2): 333-358.

Bratberg, Espen; Dahl, Svenn-Åge; Risa, Alf E. (2002): ‘The Double Burden’: Do Combinations of Career and Family Obligations Increase Sickness Absence among Women? European Sociological Review, 18(2): 233-249.

Breslin, F. Curtis; Smith, Peter (2005): Age-Related Differences in Work Injuries: A Multivariate, Population Based Study. American Journal of Industrial Medicine, 48(1): 50-56.

Brewer, Marilynn B. (1979): Ingroup Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis. Psychological Bulletin, 86(2): 307-324.

Brock, John F.; Llaneras, Robert E.; Swezey, Robert W. (1996): Older Commercial Drivers: Literature Review. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 40(18): 929-932.

Brown, Rupert J.; Turner, John C. (1981): Interpersonal and Intergroup Behaviour. In: Turner, John C.; Giles, Howard (eds.): Intergroup Behaviour. Oxford (UK): Basil Blackwell, pp. 33-65.

Brown, Sarah; Sessions, John G. (1996): The Economics of Absence: Theory and Evidence. Journal of Economic Surveys, 10(1): 23-53.

Buddlemeyer, Hielke; Wooden, Mark (2011): Transitions Out of Casual Employment: The Australian Experience. Industrial Relations, 50(1): 109-130.

*Bunderson, J. Stuart; Sutcliffe, Kathleen M. (2002): Comparing Alternative Conceptualizations of Functional Diversity on Management Teams: Process and Performance Effects. Academy of Management Journal, 45(5): 875-893.

Burdorf, Alex; Post, Wendel; Bruggeling, Ton (1996): Reliability of a Questionnaire on Sickness Absence with Specific Attention to Absence Due to Back Pain and Respiratory Complaints. Occupational and Environmental Medicine, 53(1): 58-62.

Burks, Stephen; Carpenter, Jeffrey; Goette, Lorenz (2009): Performance Pay and Worker Cooperation: Evidence from an Artefactual Field Experiment. Journal of Economic Behavior and Organization, 70(3): 458-469.

Buzzard, Richard B.; Shaw, Wendy J. (1952): An Analysis of Absence under a Scheme of Paid Sick Leave. British Journal of Industrial Medicine, 9(4): 282-295.

Byrne, Donn (1969): Attitudes and Attraction. In: Berkowitz, Leonard (ed.): Advances in Experimental Social Psychology. New York (NY): Academic Press, pp. 35-89.

Byrne, Donn (1971): The Attraction Paradigm. New York (NY): Academic Press.

Cadsby, C. Bram; Fei, Song; Tapon, Francis (2007): Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation. Academy of Management Journal, 50(2): 387-405.

*Campion, Michael A.; Papper, Ellen M.; Medsker, Gina J. (1996): Relations Between Work Team Characteristics and Effectiveness: A Replication and Extension. Personnel Psychology, 49(2): 429-452.

*Cannella, Albert A.; Park, Jong-Hun; Lee, Ho-Uk (2008): Top Management Team Functional Background Diversity and Firm Performance: Examining the Roles of Team Member Colocation and Environmental Uncertainty. Academy of Management Journal, 51(4): 768-784.

Carpenter, Jeffrey; Myers, Caitlin K. (2010): Why Volunteer? Evidence on the Role of Altruism, Image, and Incentives. Journal of Public Economics, 94(11): 911-920.

*Carpenter, Mason A.; Fredrickson, James W. (2001): Top Management Teams, Global Strategic Posture, and the Moderating Role of Uncertainty. Academy of Management Journal, 44(3): 533-545.

Chadwick-Jones, John K.; Brown, Colin A.; Nicholson, Nigel (1973): A-Type and B-Type Absence: Empirical Trends for Women Employees. Occupational Psychology, 47(1/2): 75-80.

Chadwick-Jones, John K.; Brown, Colin A.; Nicholson, Nigel; Sheppard, C. (1971): Absence Measures: Their Reliability and Stability in an Industrial Setting. Personnel Psychology, 24(3): 463-470.

Chadwick-Jones, John K.; Nicholson, Nigel; Brown, Colin A. (1982): Social Psychology of Absence. New York (NY): Praeger Publishers.

Charness, Gary (2004): Attribution and Reciprocity in an Experimental Labor Market. Journal of Labor Economics, 22(3): 665-688.

Charness, Gary; Masclet, David; Villeval, Marie C. (2013): The Dark Side of Competition for Status. Management Science, 60(1): 38-55.

Chatman, Jennifer A.; Flynn, Francis J. (2001): The Influence of Demographic Heterogeneity on the Emergence and Consequences of Cooperative Norms in Work Teams. Academy of Management Journal, 44(5): 956-974.

Che, Yeon-Koo; Yoo, Seung-Weon (2001): Optimal Incentives for Teams. American Economic Review, 91(3): 525-541.

Cheng, Grand H.-L.; Chan, Darius K.-S. (2008): Who Suffers More from Job Insecurity? A Meta-Analytic Review. Applied Psychology, 57(2): 272-303.

*Choi, Jin Nam (2007): Group Composition and Employee Creative Behavior in A Korean Electronics Company: Distinct Effects of Relational Demography and Group Diversity. Journal of Occupational and Organizational Psychology, 80(2): 213-234.

*Chowdhury, Sanjib (2005): Demographic Diversity for Building An Effective Entrepreneurial Team: Is It Important? Journal of Business Venturing, 20(6): 727-746.

Cohen, Susan G.; Bailey, Diane E. (1997): What Makes Teams Work: Group Effectiveness Research from the Shop Floor to the Executive Suite. Journal of Management, 23(3): 239-290.

Colditz, Graham A. (1999): Economic Costs of Obesity and Inactivity. Medicine and Science in Sports and Exercise, 31(11): 663-667.

*Cole, Michael S.; Bedeian, Arthur G.; Bruch, Heike (2011): Linking Leader Behavior and Leadership Consensus to Team Performance: Integrating Direct Consensus and Dispersion Models of Group Composition. Leadership Quarterly, 22(2): 383-398.

Costa, Giovanni (2003): Shift Work and Occupational Medicine: An Overview. Occupational Medicine, 53(2): 83-88.

Cox, Taylor H. (1993): Cultural Diversity in Organizations: Theory, Research & Practice. San Francisco (CA): Berrett-Koehler Publishers.

Cox, Taylor H.; Blake, Stacy (1991): Managing Cultural Diversity: Implications for Organizational Competitiveness. The Executive, 5(3): 45-56.

Cox, Taylor H.; Lobel, Sharon A.; McLeod, Poppy L. (1991): Effects of Ethnic Group Cultural Differences on Cooperative and Competitive Behaviour on a Group Task. Academy of Management Journal, 34(4): 827-847.

Croson, Rachel; Gneezy, Uri (2009): Gender Differences in Preferences. Journal of Economic Literature, 47(2): 448-474.

Dagenais, Simon; Caro, Jaime; Haldeman, Scott (2008): A Systematic Review of Low Back Pain Cost of Illness Studies in the United States and Internationally. Spine Journal, 8(1): 8-20.

De Cuyper, Nele; de Jong, Jeroen; De Witte, Hans; Isaksson, Kerstin; Rigotti, Thomas; Schalk, René (2008): Literature Review of Theory and Research on the Psychological Impact of Temporary Employment: Towards A Conceptual Model. International Journal of Management Reviews, 10(1): 25-51.

De Dreu Carsten K.; Weingart, Laurie R. (2003): Task versus Relationship Conflict, Team Performance, and Team Member Satisfaction: A Meta-Analysis. Journal of Applied Psychology, 88(4): 741-749.

De Paola, Maria (2010): Absenteeism and Peer Interaction Effects: Evidence from an Italian Public Institute. Journal of Socio-Economics, 39(3): 420-428.

De Witte, Hans (2005): Job Insecurity: Review of the International Literature on Definitions, Prevalence, Antecedents and Consequences. Journal of Industrial Psychology, 31(4): 1-6.

Deadrick, Diana L.; Stone, Dianna L. (2009): Emerging Trends in Human Resource Management Theory and Research. Human Resource Management Review, 19(2): 51-52.

DeChurch, Leslie A.; Mesmer-Magnus, Jessica R. (2010): The Cognitive Underpinnings of Effective Teamwork: A Meta-Analysis. Journal of Applied Psychology, 95(1): 32-53.

Deci, Edward L. (1971): Effects of Externally Mediated Rewards on Intrinsic Motivation. Journal of Personality and Social Psychology, 18(1): 105-115.

Deci, Edward L.; Koestner, Richard; Ryan, Richard M. (1999): A Meta-Analytic Review of Experiements Examining the Effect of Extrinsic Rewards on Intrinsic Motivation. Psychological Bulletin, 125(6): 627-668.

Deci, Edward L.; Ryan, Richard M. (1985): Intrinsic Motivation and Self-Determination in Human Behavior. New York (NY): Plenum Press.

DellaVigna, Stefano; List, John A.; Malmendier, Ulrike (2012): Testing for Altruism and Social Pressure in Charitable Giving. Quarterly Journal of Economics, 127(1): 1-56.

DerSimonian, Rebecca; Laird, Nan (1986): Meta-Analysis in Clinical Trials. Controlled Clinical Trials, 7(3): 177-188.

Devine, Dennis J. (2002): A Review and Integration of Classification Systems Relevant to Teams in Organizations. Group Dynamics: Theory, Research, and Practice, 6(4): 291-310.

Devine, Dennis J.; Clayton, Laura D.; Philips, Jennifer L.; Dunford, Benjamin B.; Melner, Sarah B. (1999): Teams in Organizations Prevalence, Characteristics, and Effectiveness. Small Group Research, 30(6): 678-711.

Dewa, Carolyn S.; Lin, Elizabeth (2000): Chronic Physical Illness, Psychiatric Disorder and Disability in the Workplace. Social Science and Medicine, 51(1): 41-50.

Dionne, Georges; Dostie, Benoit (2007): New Evidence on the Determinants of Absenteeism Using Linked Employer-Employee Data. Industrial and Labor Relations Review, 61(1): 108-120.

Dohmen, Thomas; Falk, Armin (2011): Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. American Econmic Review, 101(2): 556-590.

Dolado, Juan J.; Stucchi, Rodolfo (2008): Do Temporary Contracts Affect TFP? Evidence from Spanish Manufacturing Firms. IZA Discussion Paper No. 3832.

*Drach-Zahavy, Anat; Freund, Anat (2007): Team Effectiveness under Stress: A Structural Contingency Approach. Journal of Organizational Behavior, 28(4): 423-450.

*Drach-Zahavy, Anat; Somech, Anit (2001): Understanding Team Innovation: The Role of Team Processes and Structures. Group Dynamics: Theory, Research, and Practice, 5(2): 111-123.

*Drach-Zahavy, Anat; Somech, Anit (2002): Team Heterogeneity and Its Relationship with Team Support and Team Effectiveness. Journal of Educational Administration, 40(1): 44-66.

Drago, Robert; Garvey, Gerald T. (1998): Incentives for Helping on the Job: Theory and Evidence. Journal of Labor Economics, 16(1): 1-25.

Duijts, Saskia F.; Kant, Ijmert; Swaen, Gerard M.; Brandt, Piet A.; Zeegers, Maurice P. (2007): A Meta-Analysis of Observational Studies Identifies Predictors of Sickness Absence. Journal of Clinical Epidemiology, 60(11): 1105-1115.

Dunn, Lucia F.; Youngblood, Stuart A. (1986): Absenteeism as a Mechanism for Approaching an Optimal Labor Market Equilibrium: An Empirical Study. The Review of Economics and Statistics, 68(4): 668-674.

DWD (2014): Deutscher Wetterdienst – Weather Information Online. Available at: http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=dwdwww_start&_nfls=false
Accessed October 23 2014.

Eagly, Alice H.; Kulesa, Patrick (1997): Attitudes, Attitude Structure, and Resistance to Change: Implications for Persuasion on Environmental Issues. In Brazeman, Max H.; Messick, David M.; Tenbrunsel, Ann E.; Wade-Benzoni, Kimberly A. (eds.): Environment, Ethics, and Behavior: The Psychology of Environmental Valuation and Degradation. San Francisco (CA): New Lexington Press, pp. 122-153.

Egger, Matthias; Smith, George D.; Sterne, Jonathan A. (2001): Use and Abuse of Meta-Analysis. Clinical Medicine, 1(6): 478-484.

Ellingsen, Tore; Johannesson, Magnus (2008): Pride and Prejudice: The Human Side of Incentive Theory. American Economic Review, 98(3): 990-1008.

Elsass, Priscilla M.; Graves, Laura M. (1997): Demographic Diversity in Decision-Making Groups: The Experiences of Women and People of Color. Academy of Management Review, 22(4): 946-973.

*Ely, Robin J. (2004): A Field Study of Group Diversity, Participation in Diversity Education Programs, and Performance. Journal of Organizational Behavior, 25(6): 755-780.

Engellandt, Axel; Riphahn, Regina T. (2005): Temporary Contracts and Employee Effort. Labour Economics, 12(3): 281-299.

Erikson, Rebecca J.; Nichols, Laura; Ritter, Christian (2000): Family Influences on Absenteeism: Testing an Expanded Process Model. Journal of Vocational Behavior, 57(2): 246-272.

European Foundation for the Improvement of Living and Working Conditions (2010): Absence from Work. European Foundation for the Improvement of Living and Working Conditions. Available at:
http://www.eurofound.europa.eu/ewco/studies/tn0911039s/tn0911039s.htm
Accessed October 13 2014.

European Commission (2014a): MISSOC Comparative Tables Database. Available at: http://www.missoc.org/MISSOC/INFORMATIONBASE/COMPARATIVETABLES/MISSOCDATABASE/comparativeTableSearch.jsp
Accessed September 17 2014.

European Commission (2014b): Reducing CO2 Emissions from Heavy-Duty Vehicles. Available at: http://ec.europa.eu/clima/policies/transport/vehicles/heavy/index_en.htm
Accessed October 2nd 2014.

European Commission (2014c): Report from the Commission to the European Parliament and the Council on the State of the Union Road Transport Market.
Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014DC0222&from=EN
Accessed August 19 2014.

European Commission (2014d): Driving Time and Rest Periods. Available at:
http://ec.europa.eu/transport/modes/road/social_provisions/driving_time/index_en.htm
Accessed August 19 2014.

Eurostat (2014a): Unemployment Rate by Sex and Age Groups. Available at:
http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_rt_m&lang=en
Accessed September 17 2014.

Eurostat (2014b): Employment by Sex, Age and Economic Activity. Available at:
http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsq_egan2&lang=en
Accessed September 17 2014.

Eurostat (2014c): Labor Force Survey: Employees with a Contract of Limited Duration. Available at: http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tps00073&plugin=1
Accessed August 19 2014.

Falk, Armin (2007): Gift Exchange in the Field. Econometrica, 75(5): 1501-1511.

Falk, Armin; Ichino, Andrea (2006): Clean Evidence on Peer Effects. Journal of Labor Economics, 24(1): 39-57.

Falk, Armin; Kosfeld, Michael (2006): The Hidden Costs of Control. American Economic Review, 96(5): 1611-1630.

Fama, Eugene F. (1980): Agency Problems and the Theory of the Firm. Journal of Political Economy, 88(2): 288-307.

Farrell, Dan (1983): Exit, Voice, Loyalty, and Neglect as Responses to Job Dissatisfaction: A Multidimensional Scaling Study. Academy of Management Journal, 26(4): 596-607.

Farrell, Dan; Stamm, Carol Lee (1988): Meta-Analysis of the Correlates of Employee Absence. Human Relations, 41(3): 211-227.

Fehr, Ernst; Falk, Armin (2002): Psychological Foundations of Incentives. European Economic Review, 46(4): 687-724.

Fehr, Ernst; Gächter, Simon (2002): Do Incentive Contracts Undermine Voluntary Cooperation? Institute for Empirical Research in Economics, University of Zurich, Working Paper Series, Working Paper No. 34.

Fehr, Ernst; List, John A. (2004): The Hidden Costs and Returns of Incentives: Trust and Trustworthiness among CEOs. Journal of the European Economic Association, 2(5): 743-771.

Fekedulegn, Desta; Burchfiel, Cecil M.; Hartley, Tara A.; Andrew, Michael E.; Charles, Luenda E.; Tinney-Zara, Cathy A.; Violanti, John M. (2013): Shiftwork and Sickness Absence Among Police Officers: The BCOPS Study. Chronobiology International, 30(7): 930-941.

Fernie, Sue; Metcalf, David (1999): It's Not What You Pay it's the Way that You Pay it and that's What Gets Results: Jockeys' Pay and Performance. Labour, 13(2): 385-411.

Ferrie, Jane E.; Kivimäki, Mika; Head, Jenny; Shipley, Martin J.; Vahtera, Jussi; Marmot, Michael G. (2005): A Comparison of Self-Reported Sickness Absence with Absences Recorded in Employers’ Registers: Evidence from the Whitehall II Study. Occupational and Environmental Medicine, 62(2): 74-79.

Finkelstein, Eric; Fiebelkorn, Ian C.; Wang, Guijing (2005): The Costs of Obesity among Full-Time Employees. American Journal of Health Promotion, 20(1): 45-51.

Forrier, Anneleen; Sels, Luc (2003): Temporary Employment and Employability: Training Opportunities and Efforts of Temporary and Permanent Employees in Belgium. Work, Employment and Society, 17(4): 642-666.

Fransson, Niklas; Gärling, Tommy (1999): Environmental Concern: Conceptual Definitions, Measurement Methods, and Research Findings. Journal of Environmental Psychology, 19(4): 369-382.

Freeman, Richard B.; Kleiner, Morris M. (2005): The Last American Shoe Manufacturers: Decreasing Productivity and Increasing Profits in the Shift from Piece Rates to Continuous Flow Production. Industrial Relations, 44(2): 307-330.

Frey, Bruno S. (1997): On the Relationship between Intrinsic and Extrinsic Work Motivation. International Journal of Industrial Organization, 15(4): 427-439.

Frey, Bruno S.; Oberholzer-Gee, Felix (1997): The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out. American Economic Review, 87(4): 746-755.

Frey, Bruno S.; Oberholzer-Gee, Felix; Eichenberger, Reiner (1996): The Old Lady Visits Your Backyard: A Tale of Morals and Markets. Journal of Political Economy, 104(6): 1297-1313.

Frey; Bruno S. (1994): How Intrinsic Motivation is Crowded Out and In. Rationality and Society, 6(3): 334-352.

Frick, Bernd; Götzen, Ute; Simmons, Robert (2013): Hidden Costs of High-Performance Work Practices: Evidence from a Large German Steel Company. Industrial and Labor Relations Review, 66(1): 198-224.

Frick, Bernd; Malo, Miguel Á. (2008): Labor Market Institutions and Individual Absenteeism in the European Union: The Relative Importance of Sickness Benefit Systems and Employment Protection Legislation. Industrial Relations, 47(4): 505-529.

Gagliarducci, Stefano (2005): The Dynamics of Repeated Temporary Jobs. Labour Economics, 12(4): 429-448.

Gellatly, Ian R.; Luchak, Andrew A. (1998). Personal and Organizational Determinants of Perceived Absence Norms. Human Relations, 51(8): 1085-1102.

George, Jennifer M. (1989): Mood and Absence. Journal of Applied Psychology, 74(2): 317-324.

Georgellis, Yannis; Iossa, Elisabetta; Tabvuma, Vurain (2011): Intrinsic Motivation in the Public Sector. Journal of Public Administration Research and Theory, 21(3): 473-493.

Gibbons, Robert (1997): Incentives and Careers in Organizations. In: Kreps, David M.; Wallis, Kenneth F. (eds.): Advances in Economics and Econometrics: Theory and Application: Volume 2. Cambridge (NY): Cambridge University Press, pp. 1-37.

Gibbons, Robert (1998): Incentives in Organizations. Journal of Economic Perspectives, 12(4): 115-132.

Gibbons, Robert; Roberts, John (2013): Economic Theories of Incentives in Organizations. In: Gibbons, Robert; Roberts, John (eds.): The Handbook of Organizational Economics. Princeton (NJ): Princeton University Press, pp. 56-99.

Gneezy, Uri; Meier, Stephan; Rey-Biel, Pedro (2011): When and Why Incentives (Don't) Work to Modify Behavior. Journal of Economic Perspectives, 25(4): 191-209.

Gneezy, Uri; Rey-Biel, Pedro (2014): On the Relative Efficiency of Performance Pay and Noncontingent Incentives. Journal of the European Economic Association, 12(1): 62-72.

Gneezy, Uri; Rustichini, Aldo (2000a): Pay Enough or Don't Pay at All. Quarterly Journal of Economics, 115(3): 791-810.

Gneezy, Uri; Rustichini, Aldo (2000b): A Fine Is a Price. Journal of Legal Studies, 29(1): 1-17.

*Gong, Yaping (2006): The Impact of Subsidiary Top Management Team National Diversity on Subsidiary Performance: Knowledge and Legitimacy Perspectives. Management International Review, 46(6): 771-790.

Gordon, Michael E.; Johnson, William A. (1982): Seniority: A Review of Its Legal and Scientific Standing. Personnel Psychology, 35(2): 255-280.

Green, Colin; Kler, Parvinder; Leeves, Gareth (2010): Flexible Contract Workers in Inferior Jobs: Reappraising the Evidence. British Journal of Industrial Relations, 48(3): 605-629.

Green, Colin; Leeves, Gareth (2004): Casual Jobs and Internal Labour Markets. The Manchester School, 72(5): 658-676.

Greene, William H. (2003): Econometric Analysis. 5th ed., Upper Saddle River (NJ): Pearson Education.

*Greer, Lindred L.; Homan, Astrid C.; De Hoogh, Annebel H.; Den Hartog, Deanne. (2012): Tainted Visions: The Effect of Visionary Leader Behavior and Leader Categorization Tendencies on the Financial Performance of Ethnically Diverse Teams. Journal of Applied Psychology, 97 (1): 203-213.

Grund, Christian; Westergaard-Nielsen, Niels (2008): Age Structure of the Workforce and Firm Performance. International Journal of Manpower, 29(5): 410-422.

Guadalupe, Maria (2003): The Hidden Costs of Fixed Term Contracts: The Impact on Work Accidents. Labour Economics, 10(3): 339-357.

Güell, Maia; Petrongolo, Barbara (2001): Worker Transition from Temporary to Permanent Employment: The Spanish Case. Working Paper. Available at: http://dev3.cepr.org/meets/wkcn/4/4521/papers/petrongolo.pdf. Accessed August 19 2014.

Guest, Maya; Boggess, May M.; Duke, Janine M. (2014): Age Related Annual Crash Incidence Rate Ratios in Professional Drivers of Heavy Goods Vehicles. Transportation Research Part A: Policy and Practice, 65: 1-8.

Guinote, Ana; Fiske, Susan T. (2003): Being in the Outgroup Territory Increases Stereotypic Perceptions of Outgroups: Situational Sources of Category Activation. Group Processes and Intergroup Relations, 6(4): 323-331.

Gujarati, Damodar N.; Porter, Dawn C. (2009): Basic Econometrics. 5th ed., Boston (MA): McGraw-Hill.

Guzzo, Richard A.; Dickson. Marcus W. (1996): Teams in Organizations: Recent Research on Performance and Effectiveness. Annual Review of Psychology, 47(1): 307-338.

Hackett, Rick D. (1990): Age, Tenure, and Employee Absenteeism. Human Relations, 43(7): 601-619.

Häkkänen, Helinä; Summala, Heikki (2001): Fatal Traffic Accidents among Trailer Truck Drivers and Accident Causes as Viewed by Other Truck Drivers. Accident Analysis and Prevention, 33(2): 187-196.

*Hambrick, Donald C.; Cho, Theresa S.; Chen, Ming-Jer (1996): The Influence of Top Management Team Heterogeneity on Firms' Competitive Moves. Administrative Science Quarterly, 41(4): 659-684.

Hamilton, Barton; Nickerson, Jack; Owan, Hideo (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.

Hanisch, Kathy A.; Hulin, Charles L. (1990): Job Attitudes and Organizational Withdrawal: An Examination of Retirement and Other Voluntary Withdrawal Behaviors. Journal of Vocational Behavior, 37(1): 60-78.

Harrison, David A.; Klein, Katherine J. (2007): What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations. Academy of Management Review, 32(4): 1199-1228.

Harrison, David A.; Martocchino, Joseph J. (1998): Time for Absenteeism: A 20-Year Review of Origins, Offshoots, and Outcomes. Journal of Management, 24(3): 305-350.

Harrison, David A.; Price, Kenneth H.; Gavin, Joanne H.; Florey, Anna T. (2002): Time, Teams, and Task Performance: Changing Effects of Surface- and Deep-Level Diversity on Group Functioning. Academy of Management Journal, 45(5): 1029-1045.

Harrison, David A.; Sin, Hock-Peng (2006): What Is Diversity and How Should It Be Measured? In: Konrad, Alison M.; Prasad, Pushkala; Pringle, Judith K. (eds.): Handbook of Workplace Diversity. London (UK): Sage Publications, pp. 191-216.

Harrison, David. A.; Price, Kenneth H.; Bell, Myrtle P. (1998): Beyond Relational Demography: Time and the Effects of Surface- and Deep-Level Diversity on Work Group Cohesion. Academy of Management Journal, 41(1): 96-107.

Hatch, Nile W.; Dyer, Jeffrey H. (2004): Human Capital and Learning as a Source of Sustainable Competitive Advantage. Strategic Management Journal, 25(12): 1155-1178.

Hedges, Larry V.; Olkin, Ingram (1985): Statistical Methods for Meta-Analysis. Orlando (FL): Academic Press.

Hensing, Gunnel; Alexanderson, Kristina, Allebeck, Peter; Bjurulf, Per (1998): How to Measure Sickness Absence? Literature Review and Suggestion of Five Basic Measures. Scandinavian Journal of Work, Environment and Health, 26(2): 133-144.

*Henttonen, Kaisa; Janhonen, Minna; Johanson, Jan-Erik; Puumalainen, Kaisu (2010): The Demographic Antecedents and Performance Consequences of the Social Network Structure in Work Teams. Team Performance Management, 16(7/8): 388-412.

Higgins, Julian P.; Thompson, Simon G.; Deeks, Jonathan J.; Altman, Douglas G. (2003): Measuring Inconsistency in Meta-Analyses. BMJ: British Medical Journal, 327(7414): 557-560.

Higgins, Julian P.; Thompson, Simon G. (2002): Quantifying Heterogeneity in a Meta-Analysis. Statistics in Medicine, 21(11): 1539-1558.

Hirschman, Alberto O. (1970): Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States. Cambridge (MA): Harvard University Press.

Hitt, Michel A.; Bierman, Leonard; Shimizu, Katsuhiko; Kochhar, Rahu (2001): Direct and Moderating Effects of Human Capital on Strategy and Performance in Professional Service Firms: A Resource-Based Perspective. Academy of Management Journal, 44(1): 13-28.

Hoffman, L. Richard; Maier, Norman R. (1961): Quality and Acceptance of Problem Solutions by Members of Homogeneous and Heterogeneous Groups. Journal of Abnormal and Social Psychology, 62(2): 401-407.

Hoffman, L. Richard (1959): Homogeneity of Member Personality and Its Effect on Group Problem-Solving. Journal of Abnormal and Social Psychology, 58(1): 27-32.

Holmlund, Bertil; Storrie, Donald (2002): Temporary Work in Turbulent Times: The Swedish Experience. Economic Journal, 112(480): F245-F269.

Holmström, Bengt (1982): Moral Hazard in Teams. Bell Journal of Economics, 13(2): 324-340.

Holmström, Bengt; Milgrom, Paul (1991): Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Journal of Law, Economics, and Organization, 7(Special Issue): 24-52.

Hoque, Kim; Kirkpatrick, Ian (2003): Non-Standard Employment in the Management and Professional Workforce: Training, Consultation and Gender Implications. Work, Employment and Society, 17(4): 667-689.

Horwitz, Sujin K. (2005): The Compositional Impact of Team Diversity on Performance: Theoretical Considerations. Human Resource Development Review, 4(2): 219-245.

Horwitz, Sujin K.; Horwitz, Irwin B. (2007): The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography. Journal of Management, 33(6): 987-1015.

Houseman, Susan N. (2001): Why Employers Use Flexible Staffing Arrangements: Evidence from an Establishment Survey. Industrial and Labor Relations Review, 55(1): 149-170.

Huang, Yueng-Hsiang; Roetting, Matthias; McDevitt, Jamie R.; Melton, David; Smith, Gordon S. (2005): Feedback by Technology: Attitudes and Opinions of Truck Drivers. Transportation Research Part F: Traffic Psychology and Behaviour, 8(4/5): 277-297.

Hubbard, Thomas N. (2000): The Demand for Monitoring Technologies: The Case of Trucking. Quarterly Journal of Economics, 115(2): 533-560.

Ichino, Andrea; Maggi, Giovanni (2000): Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm. Quarterly Journal of Economics, 115(3): 1057-1090.

Ichino, Andrea; Riphahn, Regina T. (2004): Absenteeism and Employment Protection: Three Case Studies. Swedish Economic Policy Review, 11(1): 95-114.

Ichino, Andrea; Riphahn, Regina T. (2005). The Effect of Employment Protection on Worker Effort: Absenteeism During and After Probation. Journal of the European Economic Association, 3(1): 120-143.

Ichniowski, Casey; Shaw, Kathryn (2013): Insider Econometrics: Empirical Studies of How Management Matters. In: Gibbons, Robert; Roberts, John (eds.): The Handbook of Organizational Economics. Princeton (NJ): Princeton University Press, pp. 263-311.

Ichniowski, Casey; Shaw, Kathryn; Prennushi, Giovanna (1997): The Effects of Human Resource Practices on Manufacturing Performance: A Study of Steel Finishing Lines. American Economic Review, 87(3): 291-313.

Ilmarinen, Juhani E. (2001): Aging Workers. Occupational and Environmental Medicine, 58(8), 546-552.

*Jackson, Susan E.; Joshi, Aparna (2004): Diversity in Social Context: A Multi-Attribute, Multilevel Analysis of Team Diversity and Sales Performance. Journal of Organizational Behavior, 25(6): 675-702.

Jacob, Brian A.; Levitt, Steven D. (2003): Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. Quarterly Journal of Economics, 118(3): 843-877.

Jansen, Nicole W.; Kant, Ijmert; Nijhuis, Frans J.; Swaen, Gerard M.; Kristensen, Tage S. (2004): Impact of Worktime Arrangements on Work-Home Interference among Dutch Employees. Scandinavian Journal of Work, Environment & Health, 30(2): 139-148.

*Jehn, Karen A.; Bezrukova, Katerina (2004): A Field Study of Group Diversity, Work Group Context, and Performance. Journal of Organizational Behavior, 25(6): 703-729.

Jehn, Karen A.; Nothcraft, Gregory B.; Neale, Margaret A. (1997): Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups. Administrative Science Quarterly, 44(4): 741-763.

Jirjahn, Uwe (2008): On the Determinants of Shift Work and Overtime Work: Evidence From German Establishment Data. British Journal of Industrial Relations, 46(1): 133-168.

Johansson, Per; Palme, Mårten (2005): Moral Hazard and Sickness Insurance. Journal of Public Economics, 89(9): 1879-1890.

Johansson, Per; Palme, Mårten (2002): Assessing the Effect of Public Policy on Worker Absenteeism. Journal of Human Resources, 37(2): 381-409.

Johns, Gary K. (1994): How Often Were You Absent? A Review of the Use of Self-Reported Absence Data. Journal of Applied Psychology, 79(4): 574-591.

Johns, Gary K. (1997): Contemporary Research on Absence From Work: Correlates, Causes, and Consequences. International Review of Industrial and Organizational Psychology, 12: 115-174.

Jones, Derek C.; Kalmi, Panu; Kauhanen, Antti (2010): Teams, Incentive Pay, and Productive Efficiency: Evidence from a Food-Processing Plant. Industrial and Labor Relations Review, 63(4): 606-626.

Jones, Derek C.; Kato, Takao (1995): The Productivity Effects of Employee Stock-Ownership Plans and Bonuses: Evidence from Japanese Panel Data. American Economic Review, 85(3): 391-414.

*Joshi, Aparna; Liao, Hui; Jackson, Susan E. (2006): Cross-Level Effects of Workplace Diversity on Sales Performance and Pay. Academy of Management Journal, 49(3): 459-481.

Joshi, Aparna; Roh, Hyuntak (2009): The Role of Context in Work Team Diversity Research: A Meta-Analytic Review. Academy of Management Journal, 52(3): 599-627.

Kahn, Charles M.; Silva, Emilson C.; Ziliak, James P. (2001): The Brazilian Tax Collection Reform and Its Effects. Economic Journal, 111(468): 188-205.

Kaiser, Carl P. (1998): What Do We Know About Employee Absence Behavior? An Interdisciplinary Interpretation. Journal of Socio-Economics, 27(1): 79-96.

Kandel, Eugene; Lazear, Edward P. (1992): Peer Pressure and Partnerships. Journal of Political Economy, 100(4): 801-817.

*Kang, Hye-Ryun; Yang, Hee-Dong; Rowley, Chris (2006): Factors in Team Effectiveness: Cognitive and Demographic Similarities of Software Development Team Members. Human Relations, 59(12): 1681-1710.

Katzenbach, Jon R.; Smith, Douglas K. (2005): The Discipline of Teams. Harvard Business Review, 83(7): 162–171.

*Kearney, Eric; Gebert, Diether (2009): Managing Diversity and Enhancing Team Outcomes: The Promise of Transformational Leadership. Journal of Applied Psychology, 94(1): 77-89.

*Kearney, Eric; Gebert, Diether; Voelpel, Sven C. (2009): When and How Diversity Benefits Teams: The Importance of the Team Members' Need for Cognition. Academy of Management Journal, 52(3): 581-598.

*Keck, Sara L. (1997): Top Management Team Structure: Differential Effects by Environmental Context. Organizational Science, 8(2): 143-156.

Keech, M.; Scott, A. J.; Ryan, P. J. (1998): The Impact of Influenza and Influenza-like Illness on Productivity and Healthcare Resource Utilization in a Working Population. Occupational Medicine, 48(2): 85-90.

*Keller, Robert T. (1994): Technology-Information Processing Fit and the Performance of R&D Project Groups: A Test of Contingency Theory. Academy of Management Journal, 37(1): 167-179.

*Keller, Robert T. (2001): Cross-Functional Project Groups in Research and New Product Development: Diversity, Communications, Job Stress, and Outcomes. Academy of Management Journal, 44(3): 547-555.

Kessler, Ronald C.; Greenberg, Paul E.; Mickelson, Kristin D.; Meneades, Laurie M.; Wang, Philip S. (2001): The Effects of Chronic Medical Conditions on Work Loss and Work Cutback. Journal of Occupational and Environmental Medicine, 43(3): 218-225.

Kilpeläinen, Markku; Summala, Heikki (2007): Effects of Weather and Weather Forecasts on Driver Behaviour. Transportation Research Part F: Traffic Psychology and Behaviour, 10(4):288-299.

*Kirkman, Bradley L.; Tesluk, Paul E.; Rosen, Benson (2004): The Impact of Demographic Heterogeneity and Team Leader-Team Member Demographic Fit on Team Empowerment and Effectiveness. Group and Organization Management, 29(3): 334-368.

Knapp, Guido; Hartung, Joachim (2003): Improved Tests for a Random Effects Meta-Regression with a Single Covariate. Statistics in Medicine, 22(17): 2693-2710.

Knez, Marc; Simester, Duncan (2001): Firm-Wide Incentives and Mutual Monitoring at Continental Airlines. Journal of Labor Economics, 19(4): 743-772.

Knutsson, Anders (2003): Health Disorders of Shift Workers. Occupational Medicine, 53(2): 103-108.

Knutsson, Anders; Bøggild, Henrik (2010): Gastrointestinal Disorders among Shift Workers. Scandinavian Journal of Work, Environment and Health, 36(2): 85-95.

Knutsson, Anders; Goine, Hans (1998): Occupation and Unemployment Rates as Predictors of Long Term Sickness Absence in Two Swedish Counties. Social Science and Medicine, 47(1): 25-31.

Koller, Margit (1983): Health Risks Related to Shift Work. International Archives of Occupational and Environmental Health, 53(1): 59-75.

Kristensen, Kai; Juhl, Hans J.; Eskildsen, Jacob; Nielsen, Jesper; Frederiksen, Niels; Bisgaard, Carsten (2006): Determinants of Absenteeism in a Large Danish Bank. International Journal of Human Resource Management, 17(9): 1645-1658.

Kulik, Carol T.; Bainbridge, Hugh T. (2006): HR and the Line: The Distribution of HR Activities in Australian Organizations. Asia Pacific Journal of Human Resources, 44(2): 240-256.

Laaksonen, Mikko; Mastekaasa, Arne; Martikainen, Pekka; Rahkonen, Ossi; Piha, Kustaa; Lahelma, Eero (2010): Gender Differences in Sickness Absence – The Contribution of Occupation and Workplace. Scandinavian Journal of Work, Environment & Health, 36(5): 394-403.

Lacetera, Nicola; Macis, Mario (2010): Do All Material Incentives for Pro-Social Activities Backfire? The Response to Cash and Non-Cash Incentives for Blood Donations. Journal of Economic Psychology, 31(4): 738-748.

Lau, Dora. C.; Murnighan, J. Keith (1998): Demographic Diversity and Faultlines: The Compositional Dynamics of Organizational Groups. Academy of Management Review, 23(2): 325-340.

Lavy, Victor (2002): Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. Journal of Political Economy, 110(6): 1286-1317.

Lavy, Victor (2009): Performance Pay and Teachers' Effort, Productivity, and Grading Ethics. American Economic Review, 99(5): 1979-2011.

Lazear, Edward P. (2000a): The Future of Personnel Economics. Economic Journal, 110(467): F611-F639.

Lazear, Edward P. (2000b): Performance Pay and Productivity. American Economic Review, 90(5): 1346-1361.

Lazear, Edward P. (1999): Globalization and the Market for Team-Mates. Economic Journal, 109(454): 15-40.

Lazear, Edward P. (1979): Why Is There Mandatory Retirement? Journal of Political Economy, 87(6): 1261-84.

Lazear, Edward P.; Gibbs, Michael (2009): Personnel Economics in Practice. Hoboken (NJ): Wiley.

Lazear, Edward P.; Oyer, Paul (2013): Personnel Economics. In Gibbons, Robert; Roberts, John (eds.): The Handbook of Organizational Economics. Princeton (NJ): Princeton University Press, pp. 479-519.

Lazear, Edward P.; Shaw, Kathryn (2007): Personnel Economics: The Economists' View of Human Resources. Journal of Economic Perspectives, 21(4): 91-114.

Leigh, J. Paul (1995): Smoking, Self-Selection and Absenteeism. Quarterly Review of Economics and Finance, 35(4): 365-386.

Leigh, J. Paul (1986): Correlates of Absence from Work Due to Illness. Human Relations, 39(1): 81-100.

Leigh, J. Paul (1985): The Effects of Unemployment and the Business Cycle on Absenteeism. Journal of Economics and Business, 37(2): 159-170.

Levine, John M.; Resnick, Lauren B.; Higgins, E. Tory (1993): Social Foundations of Cognition. Annual Review of Psychology, 44(1): 585-612.

*Lin, Hao-Chieh; Shih, Chih-Ting (2008): How Executive SHRM System Links to Firm Performance: The Perspective of Upper Echelon and Competitive Dynamics. Journal of Management, 34(5): 853-881.

Lipsey, Mark W.; Wilson, David B. (2001): Practical Meta-Analyis. Thousand Oaks (CA): Sage.

Livanos, Ilias; Zangelidis, Alexandros (2013): Unemployment, Labor Market Flexibility, and Absenteeism: A Pan-European Study. Industrial Relations, 52(2): 492-515.

Llaneras, Robert E.; Swezey, Robert W.; Brock, John F.; Rogers, William C.; van Cott, Harold P. (1998): Enhancing the Safe Driving Performance of Older Commercial Vehicle Drivers. International Journal of Industrial Ergonomics, 22(3): 217-245.

Lount, Robert B.; Wilk, Steffanie L. (2014): Working Harder or Hardly Working? Posting Performance Eliminates Social Loafing and Promotes Social Laboring in Workgroups. Management Science, 60(5): 1098-1106.

*Lovelace, Kay; Shapiro, Debra L.; Weingart, Laurie R. (2001): Maximizing Cross-Functional New Product Teams' Innovativeness and Constraint Adherence: A Conflict Communications Perspective. Academy of Management Journal, 44(4): 779-793.

Lundborg, Petter (2007): Does Smoking Increase Sick Leave? Evidence Using Register Data on Swedish Workers. Tobacco Control, 16(2): 114-118.

Lutsey, Nicholas; Brodrick, Christie-Joy; Sperling, Daniel; Oglesby, Carollyn (2004): Heavy-Duty Truck Idling Characteristics: Results from a Nationwide Survey. Transportation Research Record, 1880(1): 29-38.

Maetzel, Andreas; Li, Linda (2002): The Economic Burden of Low Back Pain: A Review of Studies Published Between 1996 and 2001. Best Practice and Research Clinical Rheumatology, 16(1): 23-30.

Martins, Luis L.; Miliken, Frances J.; Wiesenfeld, Batia M.; Salgado, Susan R. (2003): Racioethnic Diversity and Group Members' Experiences: The Role of the Racioethnic Diversity of the Organizational Context. Group and Organization Management, 28(1): 75-106.

Martocchio, Joseph J. (1989): Age-Related Differences in Employee Absenteeism: A Meta-Analysis. Psychology and Aging, 4(4): 409-414.

Martocchio, Joseph J. (1992): The Financial Cost of Absence Decisions. Journal of Management, 18(1): 133-152.

Martocchio, Joseph J. (1994): The Effects of Absence Culture on Individual Absence. Human Relations, 47(3): 243-262.

Mas, Alexandre; Moretti, Enrico (2009): Peers at Work. American Economic Review, 99(1): 112-145.

Mastekaasa, Arne (2000): Parenthood, Gender, and Sickness Absence. Social Science and Medicine, 50(12), 1827-1842.

Mastekaasa, Arne (2005): Sickness Absence in Female- and Male-Dominated Occupations and Workplaces. Social Science and Medicine, 60(10): 2261-2272.

*Mayo, Margarita; Pastor, Juan C.; Meindl, James R. (1996): The Effects of Group Heterogeneity on the Self-Perceived Efficacy of Group Leaders. Leadership Quarterly, 7(2): 265-284.

McCall, Brian P.; Horwitz, Irwin B. (2005): Occupational Vehicular Accident Claims: A Workers' Compensation Analysis of Oregon Truck Drivers 1990-1997. Accident Analysis and Prevention, 37(4): 767-774.

McGrath, Joseph E.; Berdahl, Jennifer L.; Arrow, Holly (1995): Traits, Expectations, Culture, and Clout: The Dynamics of Diversity in Work Groups. In: Jackson, Susan E.; Ruderman, Marian N. (eds.): Diversity in Work Teams: Research Paradigms for a Changing Workplace. Washington (DC): American Psychological Association, pp. 17-45.

Mellström, Carl; Johannesson, Magnus (2008): Crowding Out in Blood Donation: Was Titmuss Right? Journal of the European Economic Association, 6(4): 845-863.

Merkus, Suzanne L.; Van Drongelen, Alwin; Holte, Kari A.; Labriola, Merete; Lund, Thomas; Van Mechelen, Willem; Van der Beek, Allard J. (2012): The Association Between Shift Work and Sick Leave: A Systematic Review. Occupational and Environmental Medicine, 69(10): 701-712.

Mertens, Antje; Gash, Vanessa; McGinnity, Frances (2007): The Cost of Flexibility at the Margin. Comparing the Wage Penalty for Fixed-Term Contracts in Germany and Spain Using Quantile Regression. Labour, 21(4/5): 637-666.

Messing, Karen; Tissot, France M.; Saurel-Cubizolles, Marie-Josèphe; Kaminski, Monique; Bourgine, Madeleine (1998): Sex as a Variable Can Be a Surrogate for Some Working Conditions: Factors Associated with Sickness Absence. Journal of Occupational and Environmental Medicine, 40(3): 250-260.

Milgrom, Paul (1988): Employment Contracts, Influence Activities, and Efficient Organization Design. Journal of Political Economy, 96(1): 42–60.

Milgrom, Paul; Roberts, John (1990): The Economics of Modern Manufacturing: Technology, Strategy, and Organization. American Economic Review, 80(3): 511-528.

Milgrom, Paul; Roberts, John (1988): An Economic Approach to Influence Activities in Organizations. American Journal of Sociology, 94(supplement): S154-S179.

*Miller, Toyah; Triana, Maria (2009): Demographic Diversity in the Boardroom: Mediators of the Board Diversity-Firm Performance Relationship. Journal of Management Studies, 46(5): 755-786.

Milliken, Frances J.; Martins, Luis L. (1996): Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups. Academy of Management Review, 21(2): 402-433.

Miners, Ian A.; Moore, Michael L.; Champoux, Joseph E.; Martocchio, Joseph J. (1995): Time-Serial Substitution Effects of Absence Control on Employee Time-Use. Human Relations, 48(3): 307-326.

*Mohammed, Susan; Nadkarni, Sucheta (2011): Temporal Diversity and Team Performance: The Moderating Role of Team Temporal Leadership. Academy of Management Journal, 54(3): 489-508.

Molinari, Noelle-Angelique M.; Ortega-Sanchez, Ismael R.; Messonnier, Mark L.; Thompson, William W.; Wortley, Pascale M.; Weintraub, Eric; Bridges, Carolyn B. (2007): The Annual Impact of Seasonal Influenza in the US: Measuring Disease Burden and Costs. Vaccine, 25(27), 5086-5096.

Morikawa, Yuko; Miura, Katsuyuki; Ishizaki, Masao; Nakagawa, Hideaki; Kido, Teruhiko; Naruse, Yuchi; Nogawa, Koji (2001): Sickness Absence and Shift Work Among Japanese Factory Workers. Journal of Human Ergology, 30(1/2): 393-398.

Nagin, Daniel S.; Rebitzer, James B.; Sanders, Seth; Taylor, Lowell J. (2002): Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment. American Economic Review, 92(4): 850-873.

*Ndofor, Hermann A.; Sirmon, David G.; He, Xiaoming (2014): Utilizing the Firm's Resources: How TMT Heterogeneity and Resulting Faultlines Affect TMT Tasks. Strategic Management Journal, 36(11): 1656-1674.

Nemeth, Charlan. J. (1986): Differential Contributions of Majority and Minority Influence. Psychological Review, 93(1): 23-32.

Ng, Thomas W.; Feldman, Daniel C. (2008): The Relationship of Age to Ten Dimensions of Job Performance. Journal of Applied Psychology, 93(2): 392-423.

Ng, Thomas W.; Feldman, Daniel C. (2013): Employee Age and Health. Journal of Vocational Behavior, 83(3): 336-345.

Nicholson, Nigel; Brown, Colin A.; Chadwick-Jones, John K. (1977): Absence from Work and Personal Characteristics. Journal of Applied Psychology, 62(3): 319-327.

Nickerson, Jack A.; Silverman, Brian S. (2003): Why Aren't All Truck Drivers Owner-Operators? Asset Ownership and the Employment Relation in Interstate For-Hire Trucking. Journal of Economics and Management Strategy, 12(1): 91-118.

*Nielsen, Bo B.; Nielsen, Sabina (2013): Top Management Team Nationality Diversity and Firm Performance: A Multilevel Study. Strategic Management Journal, 34(3): 373-382.

Norgren, Jill (2010): Ladies of Legend: The First Generation of American Women Attorneys. Journal of Supreme Court History, 35(1): 71-90.

Norrman, Jonas; Eriksson, Marie; Lindqvist, Sven (2000): Relationships between Road Slipperiness, Traffic Accident Risk and Winter Road Maintenance Activity. Climate Research, 15(3): 185-193.

Norström, Thor; Moan, Synnøve (2009): Per Capita Alcohol Consumption and Sickness Absence in Norway. European Journal of Public Health, 19(4): 383-388.

O'Reilly, Charles A.; Caldwell, David F.; Barnett, William P. (1989): Work Group Demography, Social Integration, and Turnover. Administrative Science Quarterly, 34(1): 21-37.

OECD (2014a): Employment Policies and Data - Online OECD Employment Database. Available at: http://www.oecd.org/els/emp/onlineoecdemploymentdatabase.htm#jobvac Accessed September 17 2014.

OECD (2014b): OECD Environmental Performance Reviews. Available at: https://www.scribd.com/fullscreen/244253914?access_key=key-J78rQIXmS0Nnr762iNbt&allow_share=true&escape=false&view_mode=book Accessed October 30 2014.

OECD (2014c): Labor Market Statistic Data – Incidence of Permanent Employment. Available at: http://stats.oecd.org/Index.aspx?DatasetCode=TEMP_I. Accessed August 19 2014.

OECD (2013): Protecting Jobs, Enhancing Flexibility: A New Look at Employment Protection Legislation. In: OECD Employment Outlook 2013, OECD Publishing.

Olsson, Martin (2009): Employment Protection and Sickness Absence. Labour Economics, 16(2): 208-214.

Oyer, Paul; Schaefer, Scott (2011): Personnel Economics: Hiring and Incentives. In: Ashenfelter, Orley; Card, David (eds.): Handbook of Labor Economics: Volume 4B. San Diego (CA): North Holland, pp. 1769-1823.

Paarsch, Harry J.; Shearer, Bruce (2000): Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records. International Economic Review, 41(1): 59-92.

Paoli, Pascal; Merllié, Damien (2001): Third European Survey on Working Conditions. Luxembourg: Office for Official Publications of the European Communities.

Pelled, Lisa H. (1996): Demographic Diversity, Conflict, and Work Group Outcomes: An Intervening Process Theory. Organization Science, 7(6): 615-631.

*Pelled, Lisa H.; Eisenhardt, Kathleen M.; Xin, Katherine R. (1999): Exploring the Black Box: An Analysis of Work Group Diversity, Conflict, and Performance. Administrative Science Quarterly, 44(1): 1-28.

*Peters, Linda; Karren, Ronald J. (2009): An Examination of the Roles of Trust and Functional Diversity on Virtual Team Performance Ratings. Group and Organization Management, 34(4): 479-504.

Pfeifer, Christian (2010): Work Effort During and After Employment Probation: Evidence from German Personnel Data. Journal of Economics and Statistics, 230(1): 77-91.

Polzer, Jeffrey T.; Milton, Laurie P.; Swann, William B. (2002): Capitalizing on Diversity: Interpersonal Congruence in Small Work Groups. Administrative Science Quarterly, 47(2): 296-324.

Porter, Lyman W.; Steers, Richard M. (1973): Organizational, Work, and Personal Factors in Employee Turnover and Absenteeism. Psychological Bulletin, 80(2): 151-176.

Prat, Andrea (2002): Should a Team Be Homogeneous? European Economic Review, 46(7): 1187-1207.

Prendergast, Canice (1999): The Provision of Incentives in Firms. Journal of Economic Literature, 37(1): 7-63.

Prendergast, Canice (2008): Intrinsic Motivation and Incentives. American Economic Review, 98(2): 201-205.

Puhani, Patrick A.; Sonderhoff, Katja (2010): The Effects of a Sick Pay Reform on Absence and on Health-Related Outcomes. Journal of Health Economics, 29(2): 285-302

*Qian, Cuili; Cao, Qing; Takeuchi, Riki (2013): Top Management Team Functional Diversity and Organizational Innovation in China: The Moderating Effects of Environment. Strategic Management Journal, 34(1): 110-120.

*Raver, Jana L.; Gelfand, Michele J. (2005): Beyond the Individual Victim: Linking Sexual Harassment, Team Processes, and Team Performance. Academy of Management Journal, 48(3): 387-400.

*Reagans, Ray; Zuckerman, Ezra (2001): Network, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. Organization Science, 12(4): 502-517.

*Reagans, Ray; Zuckerman, Ezra; McEvily, Bill (2004): How to Make the Team: Social Networks vs. Demography as Criteria for Designing Effective Teams. Administrative Science Quarterly, 49(1): 101-103.

Reeson, Andrew F.; Tisdell, John G. (2008): Institutions, Motivations and Public Goods: An Experimental Test of Motivational Crowding. Journal of Economic Behavior and Organization, 68(1): 273-281.

Rentsch, Joan R.; Steel, Robert P. (2003): What Does Unit-Level Absence Mean? Issues for Future Unit-Level Absence Research. Human Resource Management Review, 13(2): 185-202.

Rhodes, Susan R. (1983): Age-Related Differences in Work Attitudes and Behavior: A Review and Conceptual Analysis. Psychological Bulletin, 93(2): 328-367.

Rhodes, Susan R. (1990): Managing Employee Absenteeism. Reading (MA): Addison-Wesley.

*Richard, Orlando C.; Barnett, Tim; Dwyer, Sean; Chadwick, Ken (2004): Cultural Diversity in Management, Firm Performance and the Moderating Role of Entrepreneurial Orientation Dimensions. Academy of Management Journal, 47(2): 255-266.

*Richard, Orlando C.; Shelor, Roger M. (2002): Linking Top Management Team Age Heterogeneity to Firm Performance: Juxtaposing Two Mid-Range Theories. International Journal of Human Resource Management, 13(6): 958-974.

Riordan, Christine M. (2000): Relational Demography within Groups: Past Developments, Contradictions, and New Directions. Research in Personnel and Human Resources Management, 19: 131-174.

Riphahn, Regina T. (2004): Employment Protection and Effort among German Employees. Economics Letters, 85(3): 353-357.

Riphahn, Regina T.; Thalmaier, Anja (2001): Behavioral Effects of Probation Periods: An Analysis of Worker Absenteeism. Journal of Economics and Statistics, 221(2): 179-201.

Roberts, John (2010): Designing Incentives in Organizations. Journal of Institutional Economics, 6(1): 125-132.

Rodriguez, Daniel A.; Targa, Felipe; Belzer, Michael H. (2006): Pay Incentives and Truck Driver Safety: A Case Study. Industrial and Labor Relations Review, 59(2): 205-225.

Roetting, Matthias; Huang, Yueng-Hsiang; McDevitt, Jamie R.; Melton, David (2003): When Technology Tells You How You Drive: Truck Drivers Attitudes towards Feedback by Technology. Transportation Research Part F: Traffic Psychology and Behaviour, 6(4): 275-287.

Rosen, Sherwin (1986): Prizes and Incentives in Elimination Tournaments. American Economic Review, 76(4): 701-715.

Rosenbaum, Milton E. (1986): Comment on a Proposed Two-Stage Theory of Relationship Formation: First, Repulsion; Then, Attraction. Journal of Personality and Social Psychology, 51(6): 1171-1172.

Rosenbaum, Stephen M.; Billinger, Stephan; Stieglitz, Niels (2014): Let's Be Honest: A Review of Experimental Evidence of Honesty and Truth-Telling. Journal of Economic Psychology, 45: 181-196.

Rosenblatt, Zehava; Shapira-Lishchinsky, Orly; Shirom, Arie (2010): Absenteeism in Israeli Schoolteachers: An Organizational Ethics Perspective. Human Resource Management Review, 20(3): 247-259.

*Rousseau, Vincent; Aubé, Caroline (2010): Team Self-Managing Behaviors and Team Effectiveness: The Moderating Effect of Task Routineness. Group and Organization Management, 35(6): 751-781.

Rummel, Amy; Feinberg, Richard (1988): Cognitive Evaluation Theory: A Meta-Analytic Review of the Literature. Social Behavior and Personality, 16(2): 147-164.

Sagie, Abraham (1998): Employee Absenteeism, Organizational Commitment, and Job Satisfaction: Another Look. Journal of Vocational Behavior, 52(2): 156-171.

*Schippers, Michaéla C.; den Hartoog, Deanne N.; Koopman, Paul L.; Wienk, Janique A. (2003): Diversity and Team Outcomes: The Moderating Effects of Outcome Interdependence and Group Longevity and the Mediating Effect of Reflexivity. Journal of Organizational Behavior, 24(6): 779-802.

Schittler, Michael (2003): State-of-the-Art and Emerging Truck Engine Technologies for Optimized Performance, Emissions and Life Cycle Costs. Availabe at: http://www.osti.gov/scitech/biblio/829810
Accessed October 15 2014.

Schneider, Henry S. (2012): Agency Problems and Reputation in Expert Services: Evidence from Auto Repair Agency Problems and Reputation in Expert Services: Evidence from Auto Repair. Journal of Industrial Economics, 60(3): 406-433.

Schweitzer, Lisa; Brodrick, Christie-Joy; Spivey, Sue E. (2008): Truck Driver Environmental and Energy Attitudes: An Exploratory Analysis. Transportation Research Part D, 13(3): 141-150.

Severens, Johan L.; Mulder, Jan; Laheij, Robert J.; Verbeek, André L. (2000): Precision and Accuracy in Measuring Absence from Work as a Basis for Calculating Productivity Costs in The Netherlands. Social Science and Medicine, 51(2): 243-249.

Shapiro, Carl; Stiglitz, Joseph E. (1984): Equilibrium Unemployment as a Worker Discipline Device. American Economic Review, 74(3): 433-444.

Shaw, James B. (2004): The Development and Analysis of a Measure of Group Faultlines. Organizational Research Methods, 7(1): 66-100.

Shaw, Kathryn (2009): Insider Econometrics: A Roadmap with Stops along the Way. Labour Economics, 16(6): 607-617.

Shearer, Bruce (2004): Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment. Review of Economic Studies, 71(2): 513-534.

Sheikh, Shahbaz A. (2007): Determinants of Asset Ownership: Some Evidence from The U.S. Trucking Industry. International Journal of Finance, 19(1): 4278-4299.

Shi, Jingye; Skuterud, Mikal (2015): Gone Fishing! Reported Sickness Absenteeism and the Weather. Economic Inquiry, 53(1): 388-405.

*Shin, Shung J.; Zhou, Jing (2007): When Is Educational Specialization Heterogeneity Related to Creativity in Research and Development Teams? Transformational Leadership as a Moderator. Journal of Applied Psychology, 92(6): 1709-1721.

Slany, Corinna; Schütte, Stefanie; Chastang, Jean-Francois; Parent-Thirion, Agnes; Vermeylen, Greet; Niedhammer, Isabelle (2014): Psychosocial Work Factors and Long Sickness Absence in Europe. International Journal of Occupational and Environmental Health, 20(1): 16-25.

*Simons, Tony; Pelled, Lisa H.; Smith, Ken A. (1999): Making Use of Difference: Diversity, Debate, and Decision Comprehensiveness in Top Management Teams. Academy of Management Journal, 42(6): 662-673.

Sliwka, Dirk (2007): Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes. American Economic Review, 97(3): 999-1012.

*Smith, Ken G.; Smith, Ken A.; Olian, Judy D.; Sims, Henry P.; O'Bannon, Douglas P.; Scully, Judith A. (1994): Top Management Team Demography and Process: The Role of Social Integration and Communication. Administrative Science Quarterly, 39(3): 412-438.

*Somech, Anit (2006): The Effects of Leadership Style and Team Process on Performance and Innovation in Functionally Heterogeneous Teams. Journal of Management, 32(1): 132-157.

*Somech, Anit; Drach-Zahavy, Anat (2013): Translating Team Creativity to Innovation Implementation: The Role of Team Composition and Climate for Innovation. Journal of Management, 39(3): 684-708.

Spence, Michael (1973): Job Market Signaling. Quarterly Journal of Economics, 87(3): 355-374.

Stasser, Garold; Stewart, Dennis D.; Wittenbaum, Gwen M. (1995): Expert Roles and Information Exchange during Discussion: The Importance of Knowing Who Knows What. Journal of Experimental Social Psychology, 31(3): 244-265.

Steers, Richard M.; Rhodes, Susan R. (1978): Major Influences on Employee Attendance: A Process Model. Journal of Applied Psychology, 63(4): 391-407.

Stein, Friedrich (2015): Shift Work Design and Worker Absenteeism – Four Economic Case Studies. Dissertation: University of Paderborn. Available at: https://digital.ub.uni-paderborn.de/hs/content/titleinfo/1759456 Accessed October 01 2015.

Stewart, Greg L. (2006): A Meta-Analytic Review of Relationships between Team Design Features and Team Performance. Journal of Management, 32(1): 29-55.

Sundstrom, Eric; Mclntyre, Michael; Halfhill, Terri; Richards, Heather (2000): Work Groups: From the Hawthorne Studies to Work Teams of the 1990s and Beyond. Group Dynamics: Theory, Research, and Practice, 4(1): 44-67.

Sverke, Magnus; Hellgren, Johnny; Näswall, Katharina (2002): No Security: A Meta-Analysis and Review of Job Insecurity and Its Consequences. Journal of Occupational Health Psychology, 7(3): 242-264.

Sydsjö, Adam; Sydsjö, Gunilla; Alexanderson, Kristina (2001): Influence of Pregnancy-Related Diagnoses on Sick-Leave Data in Women Aged 16-44. Journal of Women's Health and Gender-Based Medicine, 10(7): 707-714.

Tajfel, Henri (1982): Social Identity and Intergroup Relations. Cambridge (UK): Cambridge University Press.

Tajfel, Henri (1972): Some Developments in European Social Psychology. European Journal of Social Psychology, 2(3): 307-321.

Tajfel, Henri (1969): Cognitive aspects of Prejudice. Journal of Social Issues, 25(4): 79-97.

Tajfel, Henri; Turner, John C. (1986): The Social Identity Theory of Intergroup Behavior. In: Worchel, Stephen; Austin, William G. (eds.): Psychology of Intergroup Relations. Chicago (IL): Nelson-Hall, pp. 7-24.

Tang, Shu-Hua; Hall, Vernon C. (1995): The Overjustification Effect: A Meta-Analysis. Applied Cognitive Psychology, 9(5): 365-404.

Teachman, Jay D. (1980): Analysis of Population Diversity Measures of Qualitative Variation. Sociological Methods and Research, 8(3): 341-362.

Tenhiälä, Aino; Linna, Anne; von Bonsdorff, Monika; Pentti, Jaana; Vahtera, Jussi; Kivimäki, Mika; Elovainio, Marko (2013): Organizational Justice, Sickness Absence and Employee Age. Journal of Managerial Psychology, 28(7/8): 805-825.

Thatcher, Sherry M.; Pantel, Pankaj C. (2011): Demographic Faultlines: A Meta-Analysis of the Literature. Journal of Applied Psychology, 96(6): 1119-1139.

Thøgersen, John (1994): Monetary Incentives and Environmental Concern. Effects of a Differentiated Garbage Fee. Journal of Consumer Policy, 17(4): 407-442.

Thomson, Louise; Griffiths, Amanda; Davison, Suzanne (2000): Employee Absence, Age and Tenure: A Study of Nonlinear Effects and Trivariate Models. Work and Stress, 14(1): 16-34.

Thoursie, Peter S. (2004): Reporting Sick: Are Sporting Events Contagious? Journal of Applied Econometrics, 19(6): 809-823.

Thylefors, Ingela; Persson, Olle; Hellström, Daniel (2005): Team Types, Perceived Efficiency and Team Climate in Swedish Cross-Professional Teamwork. Journal of Interprofessional Care, 19(2): 102-114.

Timmerman, Thomas A. (2000): Racial Diversity, Age Diversity, Interdependence, and Team Performance. Small Group Research, 31(5): 592-606.

Titmuss, Richard M. (1970): The Gift Relationship: From Human Blood to Social Policy. London (UK): Allen & Unwin.

Tompa, Emile; Scott-Marshall, Heather; Fang, Miao (2008): The Impact of Temporary Employment and Job Tenure on Work-Related Sickness Absence. Occupational and Environmental Medicine, 65(12): 801-807.

Tsai, Yuping; Zhou, Fangjun; Kim, InKyu (2014): The Burden of Influenza-Like Illness in the US Workforce. Occupational Medicine, 64(5): 341-347.

Tsui, Anne S.; Egan, Terri D.; O'Reilly, Charles A. (1992): Being Different: Relational Demography and Organizational Attachment. Administrative Science Quarterly, 37(4): 549-579.

Tsui, Anne S.; Gutek, Barbara A. (1999): Demographic Differences in Organizations: Current Research and Future Directions. Lanham (MD): Lexington Books.

Turner, John C. (1987): Rediscovering the Social Group: A Social-Categorization Theory. Oxford (UK): Blackwell.

*Tyran, Kristi L.; Gibson, Cristina B. (2008): Is What You See What You Get? The Relationship among Surface-Level and Deep-Level Heterogeneity Characteristics, Group Efficacy, and Team Reputation. Group and Organizational Management, 33(1): 46-76.

United States Bureau of Labor Statistics (2014): Absences from Work of Employed Full-time Wage and Salary Workers by Occupation and Industry. United States Department of Labor, Bureau of Labor Statistics. Available at: http://www.bls.gov/cps/cpsaat47.pdf Accessed October 13 2014.

*Van Der Vegt, Gerben S.; Bunderson, J. Stuart (2005): Learning and Performance in Multidisciplinary Teams: The Importance of Collective Team Identification. Academy of Management Journal, 48(3): 532-547.

*Van Knippenberg, Daan; Dawson, Jeremy F.; West, Michael A.; Homan, Astrid C. (2011): Diversity Faultlines, Shared Objectives and TMT Performance. Human Relations, 64(3): 307-336.

Van Knippenberg, Daan; De Dreu, Carsten K.; Homan, Astrid C. (2004): Work Group Diversity and Group Performance: An Integrative Model and Research Agenda. Journal of Applied Psychology, 89(6): 1008-1022.

Van Knippenberg, Daan; Schippers, Michaéla C. (2007): Work Group Diversity. Annual Review of Psychology, 58: 515-541.

Van Mierlo, Joeri; Maggetto, Gaston; van de Burgwal, Erik; Gense, Raymond (2004): Driving Style and Traffic Measures: Influence on Vehicle Emissions and Fuel Consumption. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 218(1): 43-50.

Van Poppel, Mireille N.; De Vet, Henrica C.; Koes, Bart W.; Smid, Tjabe; Bouter, Lex M. (2002): Measuring Sick Leave: A Comparison of Self-Reported Data on Sick Leave and Data from Company Records. Occupational Medicine, 52(8): 485-490.

Vernon, David; Meier, Alan (2012): Identification and Quantification of Principal-Agent Problems Affecting Energy Efficiency Investments and Use Decisions in the Trucking Industry. Energy Policy, 49: 266-273.

Virtanen, Marianna; Kivimäki, Mika; Elovainio, Marko; Virtanen, Pekka; Vahtera, Jussi (2005): Local Economy and Sickness Absence: Prospective Cohort Study. Journal of Epidemiology and Community Health, 59(11): 973-978.

Virtanen, Marianna; Kivimäki, Mika; Joensuu, Matti; Virtanen, Pekka; Elovainio, Marko; Vahtera, Jussi (2005): Temporary Employment and Health: A Review. International Journal of Epidemiology, 34(3): 610-622.

Vistnes, Jessica P. (1997): Gender Differences in Days Lost from Work Due to Illness. Industrial and Labor Relations Review, 50(2): 304-323.

Voss, Margaretha; Floderus, Birgitta; Diderichsen, Finn (2001): Changes in Sickness Absenteeism Following the Introduction of a Qualifying Day for Sickness Benefit: Findings from Sweden Post. Scandinavian Journal of Public Health, 29(3): 166-174.

Voss, Margaretha; Stark, Alexander; Alfredsson, Lars; Vingård, Eva; Josephson, Malin (2008): Comparisons of Self-Reported and Register Data on Sickness Absence among Public Employees in Sweden. Occupational and Environmental Medicine, 65(1): 61-67.

Wanous, John P.; Youtz, Margaret A. (1986): Solution Diversity and the Quality of Groups Decisions. Academy of Management Journal, 29(1): 149-159.

Watson, Warren E.; Kumar, Kamalesh; Michaelsen, Larry K. (1993): Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups. Academy of Management Journal, 36(3): 590-602.

Webber, Sheila S.; Donahue, Lisa M. (2001): Impact of Highly and Less Job-Related Diversity on Work Group Cohesion and Performance: A Meta-Analysis. Journal of Management, 27(2): 141-162.

*Wegge, Jürgen; Roth, Carla; Neubach, Barbara; Schmidt, Klaus-Helmut; Kanfer, Ruth (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.

*Wei, Li-Qun; Lau, Chung-Ming (2012): Effective Teamwork at the Top: The Evidence From China. International Journal of Human Resource Management, 23(9): 1853-1870.

Whitaker, Stuart C. (2001): The Management of Sickness Absence. Occupational and Environmental Medicine, 58(6): 420-424.

Whitener, Ellen M. (1990): Confusion of Confidence Intervals and Credibility Intervals in Meta-Analysis. Journal of Applied Psychology, 75(3): 315-321.

WHO (2014): Global Health Atlas.
Available at: http://apps.who.int/globalatlas/dataQuery/
Accessed September 17 2014.

Wiersema, Margarethe F.; Bantel, Karen A. (1992): Top Management Team Demography and Corporate Strategic Change. Academy of Management Journal, 35(1): 91-121.

Wiersma, Uco J. (1992): The Effects of Extrinsic Rewards in Intrinsic Motivation: A Meta-Analysis. Journal of Occupational and Organizational Psychology, 65(2): 101-114.

Williams, Katherine Y.; O'Reilly, Charles A. (1998) Demography and Diversity in Organizations: A Review of 40 Years of Research. Research in Organizational Behavior, 20: 77-140.

Wooldridge, Jeffrey M. (2013): Introductory Econometrics: A Modern Approach. 5th ed., South-Western Cengage Learning.

Worldbank (2014): Data – GDP (current US$). Available at:
http://data.worldbank.org/indicator/NY.GDP.MKTP.CD/countries
Accessed September 17 2014.

Xie, Jia L.; Johns, Gary K. (2000): Interactive Effects of Absence Culture Salience and Group Cohesiveness: A Multi-Level and Cross-Level Analysis of Work Absenteeism in the Chinese Context. Journal of Occupational and Organizational Psychology, 73(1): 31-52.

*Yeh, Ying-Jung; Chou, Huey-Wen (2005): Team Composition and Learning Behaviors in Cross-Functional Teams. Social Behavior and Personality: An International Journal, 33(4): 391-402.

Young, Gary J.; Beckman, Howard; Baker, Errol (2012): Financial Incentives, Professional Values and Performance: A Study of Pay-for-Performance in a Professional Organization. Journal of Organizational Behavior, 33(7): 964-983.

Zellner, Arnold (1962): An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association, 57(298): 348-368.

Zenger, Todd R.; Lawrence, Barbara S. (1989): Organizational Demography: The Differential Effects of Age and Tenure Distributions onTechnical Communication. Academy of Management Journal, 32(2): 353-376.

Zenger, Todd R.; Marshall, C.R. (2000): Determinants of Incentive Intensity in Group-Based Rewards. Academy of Management Joumal, 43(2): 149-163.

Ziebarth, Nicolas R.; Karlsson, Martin (2010): A Natural Experiment on Sick Pay Cuts, Sickness Absence, and Labor Costs. Journal of Public Economics, 94(11): 1108-1122.

Ziebarth, Nicolas R.; Karlsson, Martin (2013): The Effects of Expanding the Generosity of the Statutory Sickness Insurance System. Journal of Applied Econometrics, 29(2): 208-230.

Eidesstattliche Erklärung

Hiermit versichere ich, Konstantin Böddeker, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur und Hilfsmittel erstellt zu haben. Alle Stellen, die wörtlich oder sinngemäß veröffentlichtem oder unveröffentlichtem Schrifttum entnommen sind, habe ich als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Paderborn, den 19.05.2016 Konstantin Böddeker