Symposium on
Evidence-Based Management

Editorial Preface

by

Florian Englmaier, Guido Friebel, Gerd Muehlheusser, and Andreas Roider[*]

1 Introduction

Evidence-based management (EBM) is more than yet another buzz word of management gurus. Rather, it constitutes a major progress in economists' work on organizations, in the fields of research, teaching, and transfer of knowledge alike. Through the combination of various empirical approaches and a close interaction between theory and empirical work, EBM aims to identify the causal effects of management practices and organizational choices, thereby lending its recommendations relevance and methodological rigor.

Studying organizations and management has a long-standing tradition in economics. Much progress on these issues has been made through the contributions of, among others, Coase, Barnard, Williamson, the Carnegie school, team theory, contract theory, and more recently behavioral economics. The state of economics research on organizations and management is presented and discussed comprehensively in the recent *Handbook of Organizational Economics*, edited by Gibbons and Roberts (2013), and the economics profession has acknowledged the relevance of this research through a number of Nobel prizes. Also, the usefulness of an economic approach for understanding organizations has been brought to the public's attention by best-selling books, such as *Freakonomics* (Levitt and Dubner, 2005) or *The Org* (Fisman and Sullivan, 2014).

Over the last two decades, the study of management and organizations has opened itself to empirical analyses, thus allowing for causal inference. Today, there are numerous initiatives to bridge the gap between research and practice, for example through the use of randomized field experiments, structural econometrics, insider econometrics, and laboratory experiments.

The increasing importance of empirical analyses has been made possible by the development of new theoretical tools in incentive theory and behavioral economics as well as by econometric and experimental innovation. While traditional empirical analyses of management practices and organizational patterns have focused on providing correlational evidence, in the past two decades experimental economists and econometricians have developed new tools to identify causal effects. Moreover, in the recent past, new data sources have become available (e.g., the World Management Survey (see Bloom and Van Reenen, 2007); the managerial time use project (Bandiera, Prat, and Sadun, 2013); or, more generally, access to "big data"). Last but not least, the presence of organizational economists in leading business schools and universities, and increasing openness of organizations to collaboration with researchers, have further spurred the field.

The research program of EBM aims to apply and further develop these methods in order to better understand the mechanisms that govern economic decision-making in organizations and to properly evaluate the costs and benefits of interventions. Like evidence-based medicine, it is not content with mere speculation about the possible effects of an intervention, but seeks to generate clean evidence (e.g., from randomized experiments) to identify the mechanisms that actually cause an effect. In a quest for a deeper understanding, there is growing interaction between economists and neighboring disciplines such as psychology, sociology, and management studies.

We are pleased to bring together four papers (Englmaier and Schüßler, 2016; Bellemare, Marchand, and Shearer, 2016; Kampkötter and Sliwka, 2016; Gil and Zanarone, 2016), each of which, in a specific context, lays out the potential and methodology of EBM. Each paper has been evaluated by several anonymous referees and has benefited from their insightful comments, for which we are very grateful. We hope that readers find this symposium as interesting to read and discuss as we did in putting it together.[1]

2   Contributions to this Symposium

Englmaier and Schüßler's (2016) point of departure is the well-established observation that in many industries there are persistent productivity differences, which

---

[1] The symposium intends to spark interest in the area of EBM, and it focuses on a limited number of challenging aspects. It cannot, and does not aim to, provide a comprehensive overview of this rapidly evolving field. A good starting point for such a broader endeavor is the respective chapters in the *Handbook of Organizational Economics* (Gibbons and Roberts (eds.), 2013) or the recent survey by Bloom et al. (2014).

are rather pronounced even in well-developed countries. Building on the literature on strategic human resource management (HRM), they discuss how complementarities in HRM practices might be able to explain such productivity differences. In particular, Englmaier and Schüßler (2016) argue that behavioral phenomena and heterogeneity among workers are an important explanation for the occurrence of such complementarities in the first place. For example, suppose that the pool of potential workers consists of two groups: reciprocal workers and workers with standard preferences. In this case, Englmaier and Schüßler (2016) illustrate that two sets of firms with rather distinct sets of HRM practices might emerge: a first set of firms that hire predominantly from the set of reciprocal agents, delegate decision-making, monitor little, and provide only muted explicit incentives, and a second set of firms that hire "standard" agents, delegate little, and provide strong explicit incentives. Both of these two types of firms might be comparably successful, thereby potentially explaining observed within-industry heterogeneity. In line with the paradigm of EBM, Englmaier and Schüßler (2016) argue for a close interaction of (behavioral) theory and empirical work that aims to identify the causal links between behavioral phenomena, complementarities between HRM practices, and persistent productivity differences. For example, survey evidence on management practices can lead to the development of new theories, which in turn can be tested and refined by the help of experimental methods.

In their contribution to this symposium, Bellemare, Marchand, and Shearer (2016) argue that, for a better understanding of incentive contracts and performance schemes, combining structural estimation and experiments within the same study can be very fruitful, and they call for more empirical work along these lines. Bellemare, Marchand, and Shearer (2016) demonstrate the potential benefits of combining these two methods by means of two examples (Bellemare and Shearer, 2011, 2013). For example, Bellemare and Shearer (2011) study gift-exchange within a firm. They conduct experiments that vary the magnitude of gifts and study workers' responses. However, as they convincingly argue, from a practical perspective not all possible contracts can be tested experimentally. Hence, from their experimental data, they structurally estimate reciprocity and effort cost parameters in order to predict the profitability of hypothetical contracts. Thereby, they illustrate how, by using structural estimation, experimental findings from a given context can be extended to other (labor market) settings. In a nutshell, Bellemare, Marchand, and Shearer (2016) point out that in (field) experiments, running treatments on all parameter constellations of interest is in general not feasible, while structural estimation might not be possible with naturally occurring data. Combining the two approaches can help to overcome both types of problems. However, Bellemare, Marchand, and Shearer (2016) also highlight potential pitfalls, e.g., they caution that this approach, in general, requires selected subject pools.

Kampkötter and Sliwka (2016) also argue that the complementary use of instruments may enhance our understanding of the effects of management practices. They observe that lab experiments allow researchers to measure mechanisms with precision, but do so at the expense of a lack of external validity of the results. Field

experiments, on the other hand, allow inference on the effects of HR policies in the field, but only offer the possibility of varying a small number of policies in a given context. Using the example of subjective performance evaluations, an HR instrument that is widely used in many organizations, they show how different types of experiments, economic modeling, and the use of survey data can be combined to generate knowledge about the behavioral mechanisms through which subjective performance evaluation affects outcomes and about the causal effects of subjective performance evaluation on performance.

Kampkötter and Sliwka (2016) reveal some robust patterns. In particular, differentiation in evaluation tends to increase performance provided the interdependence between the assessed employees is not too strong. However, this may come at the expense of counterproductive behavior, especially when cooperation and teamwork are important. Furthermore, using an excessively fine-grained differentiation without objective performance information may harm performance because it shifts employees' reference points and triggers negative reciprocal reactions, which tends to outweigh the potential positive incentive effects.

Gil and Zanarone (2016) discuss evidence-based research in the context of relational contracts. Neoclassical economics assumes court enforcement to be available at nonprohibitive costs. Frequently, however, we observe that informal arrangements play an important role in facilitating efficiency-enhancing exchange, for instance, in buyer–seller or employer–employee relations. An important strand of literature on relational contracts (see, e.g., MacLeod and Malcomson, 1989; Levin, 2003) has argued that two or more parties may use the threat of dissolving an efficient relationship as a sanctioning device, which helps to provide both sides with incentives to provide efficient levels of effort.

Gil and Zanarone (2016) discuss recent developments in the literature, and suggest a framework in which the insights from the relational-contracting literature can be tested empirically. They argue that two assumptions must be relaxed for the literature to address real-world situations. Firstly, in many situations, parties are not symmetrically informed about the value of relationships. Secondly, parties may be constrained in using discretionary payments in order to share the long-term rents from cooperation. Taking these relaxed assumptions into account leads to the possibility that optimal informal contracts are not stationary. What seems as a complication at first glance, though, gives rise to interesting opportunities: For example, theories taking these limitations into account provide the opportunity to explain how contractual relationships are built and gradually evolve, and how their evolution may be subject to path dependence.

Gil and Zanarone (2016) then focus on how two of the most prominent contributions to this new literature can be brought to the data. They show that Halac's (2012) paper that allows for asymmetric information gives rise to a number of testable predictions, in particular that at the onset of a contractual relationship, the bonuses in an informal incentive contract should be higher than the ones in an optimal formal contract. Moreover, informal bonus payments and the probability of defaulting on them should increase over time, but only if the principal as the in-

formed party has bargaining power. The second paper the authors discuss in detail is Board (2011), where parties may be constrained in their ability to make transfers between them. Among the testable predictions of this model are that a buyer's loyalty to its suppliers should be higher in countries characterized by lower court quality, i.e., when contracts tend to be informal. Moreover, buyers who decide to become disloyal and switch to new suppliers should use more formal contracts in their future dealings with inside suppliers, as they can no longer rely on informal agreements to prevent holdup. The authors then discuss the available data sets that could be used to test these predictions, and they review a number of other papers tackling the challenges posed by relaxing the constraints of symmetric information about the value of a relationship and of unlimited ability to transfer long-term rents.

3 Conclusion and the Way Ahead

This symposium on EBM presents a collection of papers that are meant to exemplify a number of common themes. While all the papers are of an empirical nature, they are also well grounded in economic theory. The authors agree that a sound theoretical basis and, where possible, microeconomic modeling are a prerequisite of EBM. The presented empirical methods and data sources cover a broad range, from field and laboratory experiments to the use of survey, accounting, and personnel record data. The interest of economists in managerial and organizational questions has been spurred considerably by the availability of such data, and as a result the theories on management and organizations have been broadened and deepened. To a similar extent, economists are examining a plethora of organizational policies.

The challenges ahead do not just involve methodological aspects, in particular how to connect clean causal inference with relevance and external validity. While in development economics the literature has made use of controlled field experiments for some decades, these methods are still relatively new in the economics of management and organization. From our own and the authors' experience, we believe, however, that in many corporations and public organizations, a culture of EBM is beginning to emerge in a way comparable to that among development economists and practitioners. This opens the door for a multitude of opportunities for projects between researchers and managers, in mature and developing economies alike.

From a practical perspective, some obstacles need to be overcome. In particular, the type of research discussed in this issue requires – in general – the collaboration of firms and other organizations. While management scholars have been pushing for an evidence-based management style for some time (see, e.g., Pfeffer and Sutton, 2006; Rousseau, 2006) some stakeholders in organizations may resist such an approach. For example, they may see their position jeopardized when transparency is increased or take issue with the transfer of proprietary data to researchers. This requires that researchers be sensitive to their counterparts' perspective and concerns. Building mutual trust and making data available for analysis takes time and commitment, but it can lead to deep insights into the functioning

of management and organizations, leading to win–win situations for both managers and researchers. The Committee for Organizational Economics just established within the German Economic Association (Verein für Socialpolitik) aims to provide institutional support – and training opportunities – for researchers interested in establishing such relationships. We are convinced that in the years to come, EBM will have a deep impact in the study of organizations. We are looking forward to witnessing this development.

References

Bandiera, O., A. Prat, and R. Sadun (2013), "Managing the Family Firm: Evidence from CEOs at Work," Working Paper 19722, NBER, Cambridge (MA).

Bellemare, C., S. Marchand, and B. Shearer (2016), "Structural Estimation and Experiments: Applications to Contracting Models," *Journal of Institutional and Theoretical Economics (JITE)*, 172(2), forthcoming.

— and B. Shearer (2011), "On the Relevance and Composition of Gifts within the Firm: Evidence from Field Experiments," *International Economic Review*, 52(3), 855–882.

— and — (2013), "Multidimensional Heterogeneity and the Economic Importance of Risk and Matching: Evidence from Contractual Data and Field Experiments," *The RAND Journal of Economics*, 44(2), 361–389.

Bloom, N., R. Lemos, R. Sadun, D. Scur, and J. Van Reenen (2014), "The New Empirical Economics of Management," *Journal of the European Economic Association*, 12(4), 835–876.

— and J. Van Reenen (2007), "Measuring and Explaining Management Practices across Firms and Countries," *The Quarterly Journal of Economics*, 122(4), 1351–1408.

Board, S. (2011), "Relational Contracts and the Value of Loyalty," *The American Economic Review*, 101(7), 3349–3367.

Englmaier, F., and K. Schüßler (2016), "Complementarities of Human-Resource-Management Practices: A Case for a Behavioral-Economics Perspective," *Journal of Institutional and Theoretical Economics (JITE)*, 172(2), forthcoming.

Fisman, R., and T. Sullivan (2014), *The Org: How the Office Really Works*, John Murray, London.

Gibbons, R., and J. Roberts (eds.) (2013), *The Handbook of Organizational Economics*, Princeton University Press, Princeton (NJ).

Gil, R., and G. Zanarone (2016), "New Frontiers in the Empirical Research on Informal Contracting," *Journal of Institutional and Theoretical Economics (JITE)*, 172(2), forthcoming.

Halac, M. (2012), "Relational Contracts and the Value of Relationships," *The American Economic Review*, 102(2), 750–779.

Kampkötter, P., and D. Sliwka (2016), "The Complementary Use of Experiments and Field Data to Evaluate Management Practices: The Case of Subjective Performance Evaluations," *Journal of Institutional and Theoretical Economics (JITE)*, 172(2), forthcoming.

Levin, J. (2003), "Relational Incentive Contracts," *The American Economic Review*, 93(3), 835–857.

Levitt, S. D., and S. J. Dubner (2005), *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, William Morrow, New York.

MacLeod, W. B., and J. M. Malcomson (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57(2), 447–480.

Pfeffer, J., and R. I. Sutton (2006), "Evidence-Based Management," *Harvard Business Review*, 84(1), 1–13.

Rousseau, D. M. (2006), "Is there Such a Thing as 'Evidence-Based Management'?" *Academy of Management Review*, 31(2), 256–269.

Florian Englmaier
Department of Economics &
Organizations Research Group (ORG)
LMU Munich
Kaulbachstraße 45
80539 Munich
Germany
florian.englmaier@econ.lmu.de

Guido Friebel
Department of Management
and Microeconomics
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 4
60323 Frankfurt am Main
Germany
gfriebel@wiwi.uni-frankfurt.de

Gerd Muehlheusser
Department of Economics
University of Hamburg
Von-Melle-Park 5
20146 Hamburg
Germany
gerd.muehlheusser@wiso.uni-hamburg.de

Andreas Roider
Department of Economics
University of Regensburg
Universitätsstraße 31
93040 Regensburg
Germany
andreas.roider@ur.de

Complementarities of Human-Resource Management Practices: A Case for a Behavioral-Economics Perspective

by

Florian Englmaier and Katharina Schüßler[*]

We stress the relevance of complementarities of human-resource management (HRM) practices for explaining persistent productivity differences across organizations. We give an overview of economics as well as strategic HRM literature, document gaps, and show how insights from behavioral economics can inform the analysis. To this end, we develop a simple agency model illustrating how social preferences influence the design and consequences of incentive schemes, investigate how auxiliary HRM practices can strengthen this interaction, and discuss other behavioral subfields that are also suited to inform research on complementarities. Finally, we give an outlook on further research in this field. (JEL: D22, M50, M52)

1 Motivation

One of the most important developments in the study of firms in economics and management has been the increasing availability of firm-level micro data and the ensuing emphasis on firm heterogeneity. Newly available large and detailed data sets have provided strong evidence for enormous and persistent heterogeneity of firms (and workers) over a range of dimensions, even in narrowly defined industries. These observations are starkly at odds with theoretical predictions and are commonly referred to as *persistent productivity differences* (PPDs) across firms (see Syverson, 2011). Theory predicts that competitive forces will induce firms

to quickly adopt innovations, leaving room for short-term heterogeneity, but not for the persistent patterns that emerge in the data. Indeed, PPDs are evident even in seemingly uniform industries without room for differentiation, with high competition and with minor frictions. While these differences are quite impressive for developed countries like the U.S., where a firm at the 10th percentile of the productivity distribution generates only half of the output that a firm at the 90th percentile is able to generate with the same input (Syverson, 2004), they seem to be even more pronounced for countries like India and China, where the average total factor productivity differentials between the 10th and 90th percentiles are larger than 1 : 5 (Hsieh and Klenow, 2009).

In fact, researchers have come up with several explanations, which include, for instance, differences in input quality (e.g., Abowd et al., 2005), learning-by-doing (e.g., Benkard, 2000), or differences in management practices (e.g., Bloom et al., 2014); however, even when controlling for these facts, there still remains substantial variation (see Syverson, 2011). From the perspective of firms, increasing productivity and thus ensuring that one is (and will stay) on the upper tail of the productivity distribution can be seen as a central goal, which gives rise to the question of how that can be achieved. By drawing on the literature on strategic human-resource management (SHRM), we argue that complementary HRM practices are an important part of the answer and thus also important for understanding PPDs. Specifically, strategic management in general has been traditionally focused on how to achieve a sustained competitive advantage, which can be depicted as the antecedent of PPDs (Baron and Kreps, 1999; Lockett and Thompson, 2001). Besides that, the notion of complementarities is essentially the same as the notion of fit, which is a prominent theme in SHRM: while internal fit points to the fact that the implementation of matching practices can yield convex returns, external fit makes the same claim for taking into account external contingencies (Kepes and Delery, 2007).

In this paper, we want to build on this idea and incorporate an additional aspect into the debate on complementary effects of HRM practices as a possible mechanism for bringing about PPDs. Behavioral economics highlights additional dimensions of potential complementarities by introducing individual-level heterogeneity in preferences among employees. Assuming that firms can recruit from a heterogeneous labor force has stark consequences for complementary effects of HRM practices and can result in fundamentally different systems of practices; e.g., recruiting workers with strong social preferences, much authority, little monitoring, and relatively mute explicit incentives versus recruiting workers with no social preferences, very formalized processes with little authority, and relatively strong explicit incentives. These starkly different systems of practices may yield comparably high returns as long as they are tailored to the composition of the workforce. Thus, considerations based on behavioral economics give rise to substantially richer interactions.

We give a brief summary over different views in SHRM in section 2 before focusing on two macro-level approaches to measuring the impact of HRM prac-

tices used in organizational economics: the World Management Survey (WMS) and *insider econometrics* studies (see Ichniowski and Shaw, 2003). In section 3, we then demonstrate what behavioral economics is able to contribute by making a case for focusing more on microfoundations. We do this by developing a simple model introducing agents with social preferences into the standard principal–agent framework, relating our insights to empirical findings, and posing new questions. Building on these, in section 4, we describe what we identify as an empirical agenda and conclude.

2   Review of the Literature

This section comprises two parts: first, we aim at a brief, concise explanation of the general view on complementarities and HRM practices; then, we provide an overview on empirical methods to study complementarities of HRM practices.[1]

2.1   Concepts

2.1.1   Complementarities

As Brynjolfsson and Milgrom (2013) point out, complementarities can be thought of as a set of interrelated decisions a firm has to take in order to function properly. Assume for example a firm follows a low-cost strategy; then, subsequent decisions concerning the acquisition of and the investment in human capital depend on this strategy. It can be assumed that the decisions to keep hiring and training costs low are indeed complementary to the strategy decision. Holmstrom and Milgrom (1994, p. 973) describe the general pattern of practices being complementary when "using one more intensely increases the marginal benefit of using others more intensely".[2]

Obviously, this way of thinking can be applied to a variety of situations. As shown in the example above, in organizational economics, thinking about complementarities between organizational decisions or practices has proven useful for explaining predominant practice patterns as systems of complements that then appear together (see Brynjolfsson and Milgrom, 2013, for a concise treatment of complementarity in organizations).

---

[1] It is important to note that this article is not meant to be an exhaustive overview on the extensive literature on the World Management Survey, insider-econometric studies, or complementarities of HRM practices. We rather provide a brief summary of these literatures to act as a background for developing our main argument – that behavioral-economics insights can contribute to the study of complementarities of HRM practices – and, based on that sketch, a research agenda. Almost necessarily, in doing so we paint a subjective picture and brazenly overrepresent our own work throughout the paper. For excellent reviews see, e.g., Ichniowski and Shaw (2013), Bloom et al. (2014), or Jackson, Schuler, and Jiang (2014).

[2] Aside from this insightful and straightforward intuition, economists have defined complementarities with mathematical precision using the concept of supermodularity (Milgrom and Roberts, 1995).

2.1.2 Strategic Human Resource Management and the Influence of HRM Practices

In a similar vein but largely unnoticed by economists, scholars in SHRM have been investigating the influence of HRM practices on organizational-level outcomes in general for more than 20 years (Huselid, 1995; Becker and Gerhart, 1996; Becker and Huselid, 2006). In this still-ongoing discourse, increasingly sophisticated theoretical formulations have been proposed (e.g., Delery, 1998; Gerhart, 2007; Jiang et al., 2012).[3] In general, three different theoretical approaches addressing the link between HRM and firm performance have been proposed: the universalistic, the contingency, and the configurational approach.

In early work, a universalistic perspective was taken, which follows the proposition that there exists a relationship between the adoption of particular HRM practices and increased organizational performance (Delery and Doty, 1996). Assuming homogeneous organizations, an adoption of those "best practices" is expected to increase firm performance independently of any contingencies (Huselid, 1995). In principle, the literature building on the World Management Survey also adopts a universalistic approach.

In contrast, the contingency approach suggests that HRM practices should be aligned with important contingency factors such as labor markets, competition, or firm strategies. Organizational performance is assumed to be fostered by the interaction between the external fit to contingencies and the set of HRM practices in place (Delery and Doty, 1996). Accordingly, the use of HRM practices lacking this external fit may result in ambiguity, which in turn reduces organizational performance (Schuler and Jackson, 1987; Schuler, 1989).

Lastly, the configurational approach assumes that HRM needs complex alignment with both external and internal contingency factors such that complex and idiosyncratic sets of practices at different levels originate (Delery and Doty, 1996). The underlying assumption of this perspective is that the use of a coherent system of mutually reinforcing HRM practices has greater effects on organizational performance than the sum of the effects of individual practices (see the literature following the insider-econometrics approach, e.g., Ichniowski, Shaw, and Prennushi, 1997). Note that this notion of complementary practices and synergies is almost identical to the idea of complements in organizational economics and in line with the formal definition of supermodularity. Furthermore, the notion of complementary HRM practices has also been raised and discussed by economists (Pfeffer, 1994; Baron and Kreps, 1999; Lazear and Shaw, 2007).

2.2 Empirics

Having discussed the general view on complementarities and HRM practices, we now focus on giving a brief overview of empirical approaches to identify complementarities and their role in explaining firm productivity.

---

[3] See Jackson, Schuler, and Jiang (2014) for a discussion of this literature.

As already described in the preceding paragraphs, there is a plethora of theories of SHRM on complementary HRM practices. However, although theorists keep on emphasizing the importance of an (internal) fit of different practices, direct empirical tests remain scarce (Cappelli and Neumark, 2001; Gerhart, 2007; Kepes and Delery, 2007). Throughout this section, our main focus is on economic approaches.

Traditionally, case studies on single firms have been used to build a firm intuition about underlying mechanisms in the complementarity–productivity relation – prominent examples include cases like Lincoln Electric's business methods and incentive design (Berg and Fast, 1975) as well as specific changes like the introduction of digital imaging technologies (Autor, Levy, and Murnane, 2002) or of an enterprise resource planning (ERP) system (McAfee, 2002).

However, as this approach can be misleading because of lacking generalizability (Brynjolfsson and Milgrom, 2013), several other methods have been used to systematically study these effects. In the following, we focus on two broad approaches, namely the WMS and insider-econometrics studies.

### 2.2.1   World Management Survey

Over the last decade, the WMS, a survey run by a team of researchers around Nicholas Bloom and John Van Reenen, has led to numerous important insights to explain PPDs across firms. In particular, the WMS explores how "good" management practices can explain firm heterogeneity and focuses on monitoring, targets, and incentives, to explore the effects of management practices on firm productivity in different sectors and countries; for an overview, see Bloom et al. (2014).[4] In closely related work, these authors have documented complementarities between (what they argue constitute) "good" management practices and more general firm investment behavior, namely in IT (Bloom, Sadun, and Reenen, 2012). They document that good management in the above sense makes IT investments more profitable and show that a significant share of the productivity advantage of U.S. firms over European firms can be explained by IT usage together with "better" management.

Although some of the heterogeneity across firms can be explained by the insights provided by WMS data, a lot of unexplained variation is left. This drawback is illustrated in Figure 1, which is based on the data used in Bloom and Van Reenen (2007), and plots the residuals of a regression of the log number of sales on the log number of employees, one of the performance measures used by Bloom and Van Reenen (2007), against the average z-standardized management scores of the surveyed firms. As the observations scatter widely around the regression line, the average management score still seems to be a noisy measure, even when controlling

---

[4]   While the WMS proper is a compilation of a large number of semistructured telephone interviews, starting out as cross-sectional but recently also building up a panel component, the correlational evidence from the WMS has recently been corroborated by smaller randomized control trials (RCTs); see, e.g., Bloom et al. (2013).

Figure 1

Residual Plot of log(sales) on Average Management Scores
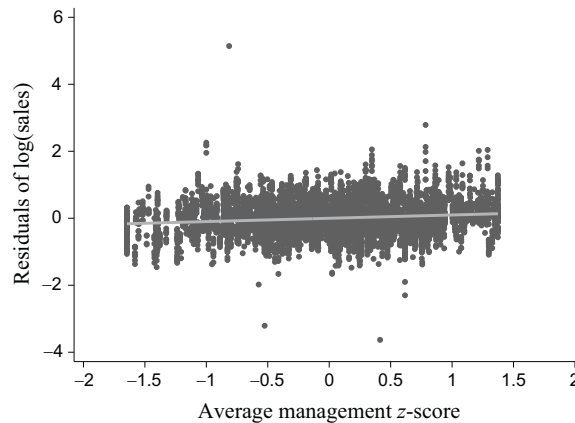
Notes: The figure plots the residuals of a regression of the log number of sales on the log number of employees, a key performance measure used by Bloom and Van Reenen (2007), against the average z-standardized management scores of the surveyed firms. Data are generously provided at http://www.worldmanagementsurvey.org/. The figure is based on our own calculations.

for firm size. Hence, to generalize from this picture, a lot of variation is left to be explained.

Concerning HRM practices, an important drawback is that the WMS focuses on a specific set of HRM practices rather than depicting the whole system of HRM practices of a firm. Furthermore, only mere correlative patterns are observed, which are not able to depict complementarities or internal fit, but only show which practices tend to be used together. Hence, even if the WMS data are helpful in explaining some of the variation in productivity, substantial amounts of PPDs remain unexplained. A particular aspect of HRM practices that is not at the center of the WMS but that might matter a lot is their complementarity. The study of these complementarities has been at the center of insider-econometrics studies, covered in the next section.

2.2.2 Insider Econometrics

Insider-econometrics studies aim at identifying the performance contribution of bundles of HRM practices more closely (Ichniowski and Shaw, 2003). This within-firm work has suggested that the specific combination of complementary HRM practices enhances productivity. Aimed at producing empirical estimates of the value of alternative HRM practices, numerous studies in this field indicate that various innovative practices are complements. The defining characteristic of a

wide range of studies that can be summarized as belonging to the insider-econo-metrics literature – see, for example, Ichniowski, Shaw, and Prennushi (1997); Lazear (2000); Hamilton, Nickerson, and Owan (2003); Shearer (2004); Bandiera, Barankay, and Rasul (2007, 2009); Wolf and Zwick (2008) – is that they use highly detailed, production-line-specific data to tie HRM practices like pay-for-perfor-mance schemes, work teams, cross-training, and routinized labor–management communication to productivity growth. In sum, these insider-econometrics stud-ies show that factors other than incentive pay are also important determinants of firm productivity.

2.3   Interim Conclusion

While both WMS and insider econometrics have been concerned with the effects of management practices on organizational performance, their focus is quite differ-ent: the former concentrates on measuring (aspects of) management quality, show-ing cross-sectional correlations with productivity (Bloom and Van Reenen, 2007), while the latter focuses on investigating within-firm data suggesting that human-resources components of management (HRM) can affect productivity in a com-plementary fashion (Ichniowski, Shaw, and Prennushi, 1997). Despite a common interest, it is important to note that both the underlying models of measurement and the theoretical rationales differ to some extent: whereas the WMS assumes ad-ditive index building with different practices causing a higher-order factor termed management quality, insider econometrics claims that those practices exhibit in-teraction effects. As Jiang et al. (2012) argue, those effects can be either negative (deadly combinations) or positive (fruitful combinations); the latter corresponds to the economic notion of complementarities. We argue that albeit WMS has been an influential and important step, from our perspective it is highly unlikely that those practices act virtually independently from one another as assumed by additive in-dexing. Moreover, we support the claim put forward by SHRM scholars, as well as by the researchers advocating the importance of complementarities in organiza-tional economics, that those firms that manage to adapt practices that are comple-mentary to (or fit their) external and internal environment are able to establish a sustained competitive advantage so that, in turn, PPDs arise.

Still, specific knowledge is fairly limited, last but not least because the tradi-tional case-study approach, as well as the correlational approach based on WMS data and insider-econometrics studies, is challenged by identification problems – unobserved heterogeneity and reverse causality being prominent ones (see Becker and Huselid, 2006, and Brynjolfsson and Milgrom, 2013, for a detailed discussion). To date, only few studies use approaches, like panel data and natural and designed experiments, that address this problem (e.g., Athey and Stern, 2002).

In addition, there is still little understanding of the microfoundations of the ef-fect of complementary HRM practices. Hence, for the remainder of this paper, our approach is to focus on two functional areas of HRM and the practices within those that have traditionally been of great interest to economists: incentive design (com-

pensation and benefits) and hiring (recruiting). Using these, we develop a simple model and use it as proof of concept to illustrate that enriching theory with behavioral concepts and sound microfoundations of individual behavior helps to better understand complementarities of HRM practices. In addition, we emphasize the idea of individual heterogeneity between workers and argue that those differences play an important role in how complementarities between practices come to life and persist.

3   What Can Behavioral-Economics Research Contribute?

After having defined our topic and briefly summarized some evidence in the previous section, we want to develop our argument that behavioral-economics research allows us to gain additional insights by highlighting an additional dimension of potential complementarities: employees' heterogeneous preferences interact in nontrivial ways with the design of incentive schemes and the choice of HRM practices more generally. This indicates a case for jointly analyzing recruiting and the organizational choices of firms.

Generally, most of the empirical literature in organizational economics so far does not focus on how recruiting and as a consequence the matching of types might interact with other HRM practices. This omission is reasonable from a neoclassical viewpoint in that these factors can to a large degree be separated from optimal incentive design. To illustrate this line of reasoning, think of worker heterogeneity: ability has been a standard dimension of heterogeneity that has been considered in the literature. Every firm wants, ceteris paribus, to hire more able workers. The firm then solves its contracting problem conditional on the characteristics of its workforce. Adding heterogeneity in risk attitudes, however, leads to a somewhat more advanced problem, as the optimal intensity of incentives is now affected by the risk attitudes of the given employee pool (see Bandiera et al., 2015). In both of these cases, ability and risk-attitude heterogeneity lead to differences in incentive heterogeneity, but do not affect the structure and nature of organizational choices in a broader sense – while there might result changes in slopes, the general structure of incentives is not changed. Thus, recruiting of workers and incentive or organizational design can be (and have been) treated largely separately.

However, in contrast to that, much research has been conducted in behavioral economics in the course of the last two decades that indicates that this clean separation of effects might not be a correct representation of reality, but that recruiting and matching of types interact with other HRM practices like incentive schemes. For instance, Delfgaauw and Dur (2007) develop a theoretical model assuming workers who are heterogeneous with respect to their intrinsic motivation, and they show how incentive schemes can help to attract motivated workers in this setting, and Jones, Willness, and Madey (2014) conduct an experiment and a field study to point out that employer signaling of corporate social performance might be a cause for selection and sorting effects of employees in firms.

Based on these insights, our goal in the remainder of this paper is to highlight how complementary effects of HRM practices and interactions with heterogeneous social preferences matter and how these can be studied. We focus on social preferences within an organizational context, and in section 3.1 we show how these preferences influence the optimal use of HRM practices by developing a simple, illustrative model introducing reciprocal preferences into the principal–agent framework based on Englmaier and Leider (2012a). We see this model as a proof of concept that behavioral-economic insights can be helpful for explaining facts and patterns in the study of HRM practices and PPDs. Of course, this is only a small piece of what behavioral-economics research has to offer, and there are numerous other areas that can be fruitfully included in the study of organizations. In section 3.2 we provide a discussion on the general mechanism behind it as well as a variety of other particularly fruitful topics.

3.1   Social Preferences: An Illustrative Model

Social preferences are a good choice to offer a proof of concept for the role behavioral-economic insights can play in studying heterogeneous preferences of agents and how they influence the optimal configuration of HRM practices. Some evidence pointing to this conclusion has been made by scholars studying these preferences.[5] First, there are numerous laboratory experiments indicating that social preferences matter in organizational settings; recently, various researchers have begun to take the question of how reciprocity matters in workplace interactions to the laboratory. Cabrales et al. (2010) find that heterogeneous social preferences, measured in standard lab tasks, are a significant predictor of choices for firms and workers in a multistage labor market experiment, and Bartling, Fehr, and Schmidt (2012) directly relate to the idea of complementary systems of HRM practices and show that both work attitude and labor market competition are causal determinants of the viability of high-performance work systems in an experimental labor market. Englmaier, Strasser, and Winter (2014) investigate one-shot labor relations in the lab and are able to show that employers take into account heterogeneity of workers in the productivity as well as in the social preference dimension. In their setting, firm profits are highest when employing reciprocal workers. Two complementary studies conducted by Englmaier and Leider (2012b,c) explore determinants for the existence of the gift-exchange mechanism both in the laboratory and in the field and show that context matters for the viability of gift exchange: in both settings, the reciprocal reaction of workers is dependent on how much the manager benefits from extra effort.

---

[5] Together with other kinds of nonstandard preferences like time preferences (present bias) and risk preferences (reference dependence), social preferences have been researched most intensely in the field of behavioral economics (DellaVigna, 2009). Fehr and Schmidt (2006) and Camerer and Weber (2013) provide extensive reviews of this line of research.

Second, there are also a few attempts to investigate this relationship in real-world settings using administrative and firm-level data, which usually use survey questions to measure reciprocal inclinations. For instance, Leuven et al. (2005) propose reciprocity as an explanation for why firms overinvest in general and specific training in comparison with standard theory predictions, and they show empirically that training investments are greater when the workforce is reciprocal, using the NIPO Post-initial Schooling Survey, a cross-sectional survey with a representative sample of the Dutch population. Dohmen et al. (2009) explore the link between reciprocal preferences and behavioral labor market outcomes and show with data from the German Socioeconomic Panel (G-SOEP) that positive reciprocity is related to high wages, high effort, and life success in general. Huang and Cappelli (2010) demonstrate that screening for "work ethic" is related to the prevalence of less monitoring, more teamwork, less employee turnover, higher wages, and higher firm productivity in U.S. census data. Englmaier, Kolaska, and Leider (2015) use personality tests in recruitment as a proxy for a reciprocal workforce and show that this is linked to wage generosity, provision of nonpecuniary incentives, teamwork, and higher firm productivity in general, using data from the UK Workplace Employment Relations Study (WERS). However, although all of these studies provide excellent starting points for future research, they all build on cross-sectional data and cannot identify causal effects.

Summing up, there is ample evidence from the lab as well as the field indicating that social preferences matter in organizational settings. Hence, in the remainder of the section, we develop a simple agency model incorporating reciprocal inclinations on the side of the agent to illustrate how behavioral economics can inform the analysis.

3.1.1 Base Model

Following Englmaier and Leider (2012a), we consider a simple binary principal–agent framework where both the principal and the agent are risk-neutral. There are two states of the world, which are characterized by two possible outcomes, q_1 and q_2, with $q_1 < q_2$ and $q_i > 0 \, \forall i \in \{1,2\}$. The agent can choose between two actions, a_L and a_H, with related costs $c(a_L) < c(a_H)$. The probabilities of the principal's outcomes are conditional on the action taken by the agent: $\Pr(q_2|a_j) = \pi_j \, \forall j \in \{L,H\}$, where $\pi_H > \pi_L$, which implies that the higher outcome is a better signal for high effort. Hence, the principal's expected return is given by $ER(a_j) = (1-\pi_j)q_1 + \pi_j q_2 \, \forall j \in \{L,H\}$.

We assume that effort is not contractible; hence, wages can only be paid conditional on outcomes. Thus, the principal offers the following contract to the agent: $(w,b,\hat{a})$, where w denotes a secure wage payment (a salary) for the agent in every state of the world, b an additional bonus that is paid in the case that outcome q_2 is realized, and $\hat{a}$ represents a nonbinding request for an action.[6] This request re-

---

[6] When developing the model in the following subsections, we always discuss optimal wages, as is common in agency models. As the model is formulated in utility terms,

flects the performance level expected by the firm, which is known to the worker (e.g., through job descriptions or a code of conduct) and serves to fix the agent's beliefs about the principal's intended generosity; see the discussion below and in Englmaier and Leider (2012a).

The agent has an inherent concern for reciprocity, η, with $\eta \in [0, +\infty)$. To focus the analysis, we assume here that this type η is publicly observable.[7] His utility from taking action a_j when $\hat{a}$ is requested is given by

$$U(a_j, \hat{a}) = (1 - \pi_j)w + \pi_j(w + b) - c(a_j)$$
$$+ \eta[(1 - \hat{\pi})w + \hat{\pi}(w + b) - c(\hat{a}) - \bar{u}]ER(a_j).$$

As is evident from the utility function, a reciprocal agent ($\eta > 0$) does not only derive utility from the wage payment as is common in agency models (first part of utility function), but also from internalizing the principal's welfare (third part). This part of the function represents the case that reciprocal utility is "triggered" when the proposed wage scheme provides the agent with a rent in excess of his outside option $\bar{u}$. In the simple case of our model, $\bar{u}$ is assumed to be fixed exogenously. For simplification, we also assume that $\bar{u} = 0$ and $c(a_L) = 0$, and define $c(a_H) = c$.

Benchmark Case.   Trivially, when assuming that effort is contractible, the first-best solution is implemented by any wage scheme compensating the agent for his effort costs c. Furthermore, without further restrictions, it can easily be shown that the principal can induce a risk-neutral agent to exert high effort for first-best costs, even when effort is not contractible.[8]

The properties of the optimal contract are summarized in the following proposition.

PROPOSITION 1 *Under hidden action, the risk-neutral principal proposes the following contract to the risk-neutral agent:*

$$w = -\frac{\pi_L}{\Delta\pi}c < 0, \quad b = \frac{c}{\Delta\pi}, \quad and \quad \hat{a} = a_H.$$

however, note that these are to be interpreted broadly and might include multiple ways of transferring utility to employees. A discussion of this is provided in section 3.2.1.

[7] If we assume instead that firms have access to a noisy signal on η upon hiring, there are two types of errors that might arise: either an in fact suitable (say reciprocal) applicant is mistaken to be nonreciprocal and is not hired, or a nonreciprocal applicant is mistaken to be a reciprocal type and is hired by a firm that has its HR systems geared towards employing reciprocal types. If, because the employment relationship is long-term, the worker continues to mimic a reciprocal type, our conclusions are not affected. If not and the worker starts to take advantage of the low-powered incentive environment, additional measures like having an intense probation period seem particularly important in such reciprocity-based settings. Even if these do not work, as long as the initial signal upon recruiting is precise, the firm presumably has to be able to live with few "rotten apples."

[8] See appendix A.1 for the exposition of the problem and a derivation of the optimal contract.

In doing so, she can induce the agent to exert a_H at first-best implementation cost $B^{FB}(a_H) = c$, leaving the agent with no rent. This holds for every $\eta \in [0, +\infty)$.

The optimal contract punishes the agent if the low outcome is realized and rewards him if the good outcome is realized. The intuitive explanation for this result is straightforward: since effort is not contractible, the principal has to align the agent's interests to hers. Recall that the agent is risk-neutral; thus, the principal can just structure incentives steeply and transfer risk to the agent without having to compensate him for that. Note that, just as in Englmaier and Leider (2012a), reciprocity does not have an effect on the optimal contract in the first best; the "standard contract" prevails.

Limited Liability.   Now, consider the case where the agent is wealth-constrained and thus has to receive a minimal wage, which cannot be negative, in every state of the world; more formally, consider $w \geq \underline{w} \geq 0$. With an additional limited-liability constraint (Innes, 1990) like this, the principal then faces a trade-off between rent extraction and incentive provision when designing the optimal contract.

The properties of the solution to this problem are summarized in the following proposition (see footnote 8).

PROPOSITION 2 *Under hidden action with limited liability ($w \geq \underline{w} \geq 0$), the risk-neutral principal proposes the following contract to the risk-neutral agent:*

$$w_1 = \underline{w}, \quad b = \frac{c}{\Delta\pi} - \frac{\eta\Delta ER}{\Delta\pi + \eta\Delta ER\pi_H}\left(\underline{w} + \frac{\pi_L}{\Delta\pi}c\right), \quad and \quad \hat{a} = a_H.$$

In doing so, she can induce the agent to exert a_H, but only at implementation cost

$$B_\eta^{SB}(a_H) = c + \frac{\Delta\pi}{\Delta\pi + \eta\Delta ER\pi_H}\left(\underline{w} + \frac{\pi_L}{\Delta\pi}c\right).$$

The agent's utility is then

$$U(a_H) = (1 + \eta ER_H)\left[\frac{\Delta\pi}{\Delta\pi + \eta\Delta ER\pi_H}\left(\underline{w} + \frac{\pi_L}{\Delta\pi}c\right)\right].$$

First of all, note that this collapses to the standard case if $\eta = 0$. Hence, the standard case is nested in our model. Comparing the optimal contract for selfish and for reciprocal agents reveals that $B_{\eta>0}^{SB}(a_H) < B_{\eta=0}^{SB}(a_H)$; i.e., the principal can implement a_H at lower cost with a reciprocal agent. In both cases, the principal has to pay a rent due to wealth constraints of the agent; however, while this rent is "lost" on a selfish agent, it serves as an additional incentive for a reciprocal agent. This is also reflected in the bonus payment b, which is lower for reciprocal agents, as the (nonzero term comprising the) limited liability rent, weighted by a term reflecting reciprocity η and the effectiveness of reciprocal behavior (ΔER), can be subtracted. The intuition for the smaller expected wage bill in this case is straightforward: the wage differential for the reciprocal agent can be smaller

because the positive rent provides an additional motivation to the reciprocal agent. More formally, $b_{\eta>0} < b_{\eta=0}$.

The three cases of the exemplary first-best contract, the second-best contract for a selfish agent, and the second-best contract for a reciprocal agent are illustrated for comparison in Figure 2.

Figure 2
Illustration of the Wage Schemes
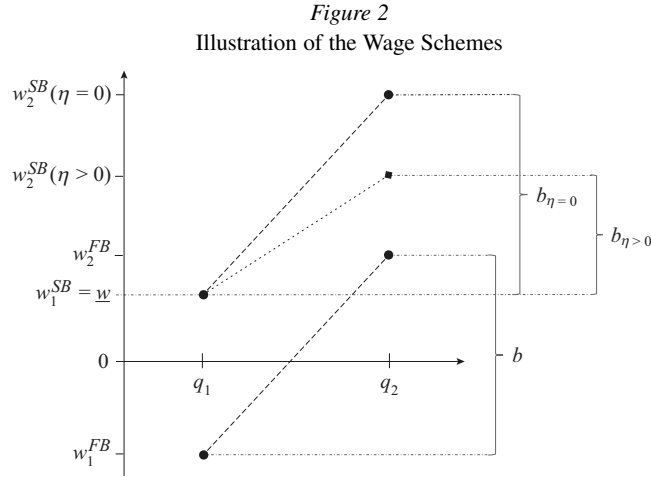
Notes: The figure depicts the exemplary first-best contract (lower dashed line), the second-best contract for a selfish agent (parallel dashed line above the first-best line) and the second-best contract for a reciprocal agent (dotted line).

Comparative Statics.   To see what happens to the wage differential if the principal is faced with a more reciprocal agent, we take the derivative of the optimal b with respect to η:

$$\frac{\partial b}{\partial \eta} = -\frac{\Delta\pi\Delta ER}{(\Delta\pi + \eta\Delta ER\pi_H)^2}\left(\underline{w} + \frac{\pi_L}{\Delta\pi}c\right) < 0.$$

Hence, the principal can offer a lower b to a more reciprocal agent, which is a rather intuitive result.[9] Subsequently, the expected wage bill for employing the agent also decreases in η. Considering the limit case $\eta \to \infty$, we can show that for an infinitely reciprocal agent, the principal could even induce high effort a_H for first-best cost:

$$\lim_{\eta\to\infty} B_{\eta}^{SB}(a_H) = \lim_{\eta\to\infty}\left[c + \frac{\Delta\pi}{\Delta\pi + \eta\Delta ER\pi_H}\left(\underline{w} + \frac{\pi_L}{\Delta\pi}c\right)\right] = c,$$

---

[9] Also, since the increase in the principal's expected revenue due to choosing a_H rather than a_L, ΔER, is always multiplied with η in b, the same holds true if the value of effort increases for the principal.

which equals $B^{FB}(a_H)$. The above discussion is summarized in the following proposition.

PROPOSITION 3 *b is decreasing in η and ΔER. As a consequence, the expected wage bill $B_\eta^{SB}(a_H)$ is also decreasing in these factors.*

Interim Conclusions.   To sum up, in a hidden-action principal–agent framework with limited liability, we can show that the principal is able to reduce the incentive intensity and the wage bill when contracting with a more reciprocal agent.

Note that even in this simple model, it is possible to immediately see that there are different strategies for firms to combine complementary HRM practices in hiring and incentive design: if a firm manages to attract a reciprocal workforce, it can implement several reciprocity-based practices (e.g., little explicit incentives, generous treatment, job security, little monitoring), while if it attracts nonreciprocal workers, the system optimally geared to this looks starkly different (e.g., internal competition, steep incentives, close performance monitoring). While these two firms, due to the use of complementary practices, look starkly different (some might say one appears well managed and the other badly managed), they might be relatively comparable in profitability.

3.1.2   Other HRM Practices in the Model Framework

In their seminal textbook, Milgrom and Roberts (1992) state that "important features of many organizations can be best understood in terms of deliberate attempts to change the preferences of individual participants [. . .]. As a result, organizationally desired behavior becomes more likely." They go on to argue that this "is clearly an element of leadership [. . .] and it has much to do with practices of organizing semipermanent groups of workers and encouraging them to interact socially as well as at work" (p. 42). While Milgrom and Roberts then continue down the neoclassical route, positing selfish, rational actors, with the progress made by behavioral-economic research we can begin to more carefully study the content of their statement. In this spirit, we include the possibility for HRM practices to influence reciprocal behavior in our model by modifying the reciprocity parameter η.

Without tapping into the vast nature-versus-nurture debate in detail, it can be stated that the assumption of stability of preferences in economics and of personality traits in psychology has been heavily challenged in recent years (Woods et al., 2013). Albeit the trait approach to personality has been the prevalent paradigm for the last eighty years, it is being criticized by scholars putting characteristics of the situation at center stage (e.g., Mischel and Shoda, 1995). In economics, the assumption of stable preferences has been scrutinized as well (see, e.g., Bowles, 1998). In light of all those issues, we argue that preferences that matter most in the work context may at least to a degree be subject to change by HRM practices in a firm. For instance, specific aspects of work design, like fostering intense teamwork, may affect preferences – in particular, strengthen preexisting social preferences.

When conceptualizing how one could incorporate HRM practices in the relatively general model framework outlined above, one can think of several different ways to do so, each with different theoretical and empirical implications. Here, we focus on ways of influencing the employee's reciprocal motivation, η, by HRM practices like team building, empowerment, or delegation of decision rights. In a similar vein, Pfeffer (2007) stresses that there are certain types of high-commitment HRM practices that serve to activate reciprocal motivation. These include investments in training, information sharing, or decentralized decision making as signals of trust. As a consequence, we have to think of η as consisting of two parts – the agent's inherent concern for reciprocity, η_1, as well as a second, malleable part that can be influenced by HRM, η_2. Depending on how these two factors interact, different predictions can be derived from the model.

To fix ideas, consider a multiplicative model: $\eta = \eta_1 \times \eta_2$.[10] For simplification, we assume that the agents' preference types are known. We substitute η in Proposition 2 and derive comparative statics by taking the derivative of the optimal bonus payment:

$$\frac{\partial b}{\partial \eta_2} = -\frac{\eta_1 \Delta \pi \Delta ER}{(\Delta \pi + \eta_1 \eta_2 \Delta ER \pi_H)^2}\left(\underline{w} + \frac{\pi_L}{\Delta \pi}c\right) < 0.$$

Here, the interplay of the two reciprocal factors is complex, as can be seen by considering the second and cross derivatives:

$$\frac{\partial^2 b}{\partial \eta_2^2} = \frac{2\eta_1^2 \Delta \pi (\Delta ER)^2 \pi_H}{(\Delta \pi + \eta_1 \eta_2 \Delta ER \pi_H)^3}\left(\underline{w} + \frac{\pi_L}{\Delta \pi}c\right) > 0,$$

$$\frac{\partial^2 b}{\partial \eta_1 \eta_2} = -\frac{\Delta \pi \Delta ER(\Delta \pi - \eta_1 \eta_2 \Delta ER \pi_H)}{(\Delta \pi + \eta_1 \eta_2 \Delta ER \pi_H)^3}\left(\underline{w} + \frac{\pi_L}{\Delta \pi}c\right) \lessgtr 0.$$

While the former is globally positive, the latter is only positive for a threshold value of $\eta_1 \eta_2 > \Delta \pi / \Delta ER \pi_H$, i.e., for a sufficient amount of reciprocal motivation already present. This hints at a separation of cases – below the threshold, investments in η_2 will hardly reduce the bonus payment necessary and are thus not profitable for the firm, while investments for levels above the threshold pay off. To illustrate this relationship, consider the situation that the firm has the opportunity to implement a training program that is known to increase η_2 from 1 to 1.5, but that has implementation costs of $1,000. Note that the investment decision cannot be made without considering the level of η_1: when it is small, the optimal bonus payment will only decrease by a small amount that does not exceed $1,000. For a large enough η_1, however, the increase of η_2 by 0.5 will result in a bonus decrease that exceeds $1,000 – in the former case, the firm optimally does not invest, while in the latter case, it does, although the effectiveness of the training itself has not changed. In this way, a higher level of η_1 increases marginal returns from investing in η_2, which

---

[10] See appendix A.2 for an additive modeling, $\eta = \eta_1 + \eta_2$, and the discussion of its implications.

Figure 3
Dependence of Optimal Bonus on η in Multiplicative Model
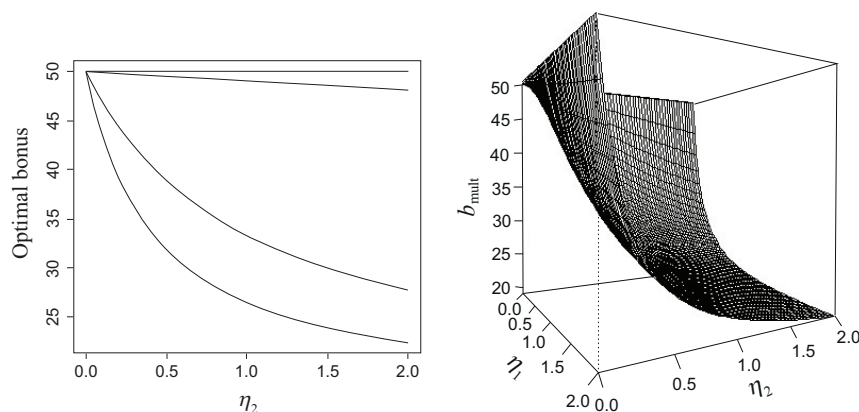
Notes: The figure depicts the optimal bonus as a function of η_1 and η_2, the parameters capturing the innate reciprocal inclination of a worker and the HRM policies targeted at increasing this reciprocal inclination towards the firm, for a multiplicative formulation $\eta = \eta_1 \times \eta_2$.

reflects the complementarity definition given above. This relationship is also depicted graphically in Figure 3.

It is evident that the firm can foster reciprocal reactions of its employees by investing in practices that increase η_2, as long as the employee has any inherent concern for reciprocity at all ($\eta_1 > 0$). Again, it is important to note that the strength of this influence and hence the decrease of the necessary bonus payment depend on the level of the agent's inherent concern for reciprocity. In other words, η_1 and η_2 are complements and thus interdependent.

Implications.   Modeling employees' preferences and preference-enhancing HRM policies as complements makes thinking about these investments more complicated (and interesting). When the ex ante level of overall reciprocal motivation is low, either because the workforce selected has a low concern for reciprocity or because there are no other investments in the form of HRM practices, additional investment does not lead to significant gains in lower bonus payments for the firm and is hence hardly profitable. In contrast, when ex ante reciprocal motivation is already above a certain threshold, gains from increasing investment in either dimension are larger. Phrased differently, as η_1 is related to hiring practices used in personnel selection, these practices are complementary to practices that directly increase η_2. As an application, one could think of using a selection process focused on screening for reciprocal agents and utilizing work design to influence reciprocity, for instance, by granting high discretion (see Pfeffer, 2007). This is then optimally complemented by a flat, generous incentive scheme.

3.1.3   Conclusions from the Model

To conclude and sum up the insights from modeling a simple contracting situation with reciprocally motivated agents, we have seen that the composition of the workforce and the effect of HRM practices are interdependent. In this respect, the design of incentive schemes and the effects from selection and matching cannot be treated separately. This in turn suggests that those practices are indeed complements yielding potential convex returns in organizational performance and, ultimately, explanatory power for disaggregating PPDs.

Before going on by broadening our focus, we now discuss in what way this model (and behavioral contracting models in general) can yield testable predictions and thus help in exploring complementarities of HRM practices. Of course, on incorporating individual preferences, the need to observe both these and individual-level outcomes is evident, something that is hard to observe in field settings.[11] However, in our opinion, this is an excellent example of a case where there is a clear role for laboratory and field experiments that make it possible to test specific mechanisms in a clean manner and to measure individual preferences using behavioral outcomes, which matters especially when individual heterogeneity comes into play (see Cohn, Fehr, and Goette, 2015, for an argument along these lines). This then, in the spirit of a feedback loop, can inform larger-scale empirical approaches using administrative data.

3.2   Other Aspects

As already stated right at the beginning of this section and hinted at in the last paragraphs, we use social preferences of workers as an illustrative example to support our case that behavioral economics has much to contribute when it comes to complementarities of HRM practices. However, of course social preferences are not the only behavioral aspect of potential relevance in this field. Possible topics for future research incorporating complementarities of HRM practices will involve further areas of behavioral-economics research. Besides social preferences, which have probably been studied most extensively, there are also various other concepts yielding insights that are likely to matter in this setting. We aim at demonstrating this point by first highlighting that the general mechanisms we demonstrated are not confined to social preferences, but rather appear to be more general – at least when utilizing a broader interpretation of the model – and then giving a brief overview of research in other areas in behavioral economics that we perceive as relevant.

---

[11] However, economic researchers are beginning to incorporate measures for individual preferences like reciprocity, risk attitudes, or patience in large administrative data sets like the G-SOEP (see, e.g., Dohmen et al., 2009, 2011), which makes it possible to investigate related questions using field data.

3.2.1 Broader Interpretations of the Model

In the following, we consider some ways in which our framework can be thought of in a more general way, indicating that we are hinting at a more general pattern. There are two aspects we want to stress: first, our results can be compared to arguments from identity economics (Akerlof and Kranton, 2000, 2005). Second, as already noted above, despite the fact that the standard agency model incorporates monetary benefits and we also frame our model in those terms, incentives do not have to be defined as being purely monetary.

Identity.   Although we define the nonstandard motivation of agents in our model as the agent being reciprocally motivated, many of our arguments are parallel to arguments in the identity literature. For instance, Akerlof and Kranton (2005) also explore the role of identity as supplement to monetary compensation by incorporating identity utility that depends on the deviation from norms and ideals. They derive a result that is similar to ours: being an insider leads to a reduced wage differential. When relating their model to the workplace, the authors explicitly state that the dichotomy of insiders and outsiders can be thought of as the dichotomy of extrinsic and intrinsic motivation in management. Here, modern approaches like investments in corporate culture aim at aligning the goals of workers with those of the firm, essentially trying to shift workers from being outsiders to being insiders.

In a more general sense, our formulation – based on Englmaier and Leider (2012a) – like the identity formulation of Akerlof and Kranton (2005), the model incorporating trust by Sliwka (2007), or the theoretical formulations incorporating motivated agents (e.g., Besley and Ghatak, 2005; Cassar, 2014), can be thought of as the utility representation of workers internalizing the firm's goals. In this respect, the aspect of motivation that is influenced by HRM practices, η_2 in our model, can be reinterpreted as investment in category affiliation, or, exactly as in the identity framework, as investment in goal alignment. The only differences between the models are then in the ways of thinking about the underlying mechanism and the underlying inherent motivation of the agents. Again, differentiating between the explanations and carving out the underlying mechanism are ultimately empirical tasks with a clear-cut role for lab and field experimental investigation.

Benefits.   As already briefly noted at the beginning of section 3, narrowly interpreted, the model (as well as the standard agency model) describes a situation in which a monetary payment, the wage, is exchanged for effort exertion by the agent. However, interpreting the model in a broader sense is possible. All the formulations are essentially set up in utility terms; hence, one could also think about the bonus as any benefit for the agent generating a utility rent and about effort exertion as everything that is beneficial for the company, which includes, for instance, retention of the employee.

Again, this way of generalizing the model has numerous empirical implications. Naturally, thinking about the problem in this generalized way means that it loses

structure, but also again highlights the role of experiments in identifying which benefits can substitute for money and which actions of the agent can substitute for effort – and, ultimately, to clarify the intricate relationship between different types of incentives and benefits (see Oyer, 2008).

3.2.2   Other Areas of Study from Behavioral Economics that Are of Particular Relevance

Besides the aspects of social preferences discussed above, there are also other interesting areas within the field of behavioral economics that might prove useful for thinking about HRM practices and possible complementarities.

First, there are an increasing number of papers focusing on different types of nonmonetary incentives in work relations. This includes research on a variety of different aspects, for example, on the broad area of "respect" (Ellingsen and Johannesson, 2007). Here, some authors have focused on exploring the role of spontaneous recognition in employee performance (e.g., Chen et al., 2010; Bradler et al., 2013; Hoogveld and Zubanov, 2014), which is proposed to work through conformity to a performance norm (Bernheim, 1994). Similarly, status incentives in general, i.e., nonpecuniary rewards for good performance like medals, prizes, awards, or job titles, have gained some interest in recent research and have been examined both theoretically (e.g., Besley and Ghatak, 2008; Auriol and Renault, 2008) and empirically (e.g., Ashraf, Bandiera, and Lee, 2014; Charness, Masclet, and Villeval, 2014; Kosfeld and Neckermann, 2011; Gubler, Larkin, and Pierce, 2013). Proposed reasons for these kinds of incentives to work are that their value is based on scarcity and the human tendency to crave social status and recognition (Besley and Ghatak, 2008). Related to that and the literature on tournaments, several researchers have also been exploring the role of relative rank as a motivator (e.g., Azmat and Iriberri, 2010; Bandiera, Larcinese, and Rasul, 2015; Barankay, 2011a,b; Blanes i Vidal and Nossol, 2011). Other topics include job characteristics like perceived meaningfulness of tasks (e.g., Ariely, Kamenica, and Prelec, 2008; Grant, 2008) or (lack of) discretion (Falk and Kosfeld, 2006; Dickinson and Villeval, 2008).

Summing up, all of these studies find that there are several other ways for principals to motivate their workers than pure incentive pay. These "new" HRM tools are of course very interesting in themselves, but so far, this area of research is mostly composed of studies that look at one tool in isolation without incorporating a whole system of practices. As with evidence on social preferences, the scope for exploring complementarities between those practices then is very limited. From our perspective, more comprehensive studies incorporating whole systems of HRM practices are needed to address possible complementarities.[12]

In addition, some researchers have also begun to look into personality traits – mostly the "Big Five" (agreeableness, conscientiousness, neuroticism, extrover-

---

[12] Recent exceptions include Bartling, Fehr, and Schmidt (2012) and Kosfeld, Neckermann, and Yang (2014).

sion, and openness) – and find that these also interact with outcome dimensions like inherent motivation and productivity (e.g., Segal, 2012). As with heterogeneous social preferences, firms' optimal systems of HRM practices are likely to depend crucially on workforce composition with respect to these traits.

Furthermore, within behavioral economics in general, there are a few other topics that we think of as promising. There is an important body of research on time-inconsistent preferences of agents that lead to self-control problems and procrastination (O'Donoghue and Rabin, 1999a,b). Here, one major insight is that sophisticates, i.e., those agents who know about their time inconsistency and seek to avoid it, might actively search for commitment devices, and firms that know about the problem of time inconsistency might in turn desire to employ sophisticates. As a consequence, offering commitment devices might be one strategy for achieving this. Examples include up-or-out schemes, conditional training or rotation schemes, and gym memberships,[13] which can all be interpreted in this way (see Englmaier, Fahn, and Schwarz, 2015, for an attempt to capture this idea theoretically within a framework of long-term employment contracting). In addition, nonstandard beliefs like overoptimism and overconfidence might also matter when thinking about interrelations between HRM practices (see, e.g., Larkin and Leider, 2012; Sautmann, 2013). However, note that in this area, as well as in other subfields of behavioral-economics research, insights almost exclusively stem from experiments conducted in laboratory settings, with few firm data available so far. Here, perhaps even more than in the area of social preferences, it is crucial to push for comprehensive multilevel field data to understand the prevalence and patterns of practices in the field.

3.3 Interim Conclusion

Throughout this section, we have made the point that enriching theoretical models with behavioral aspects has several important implications for the study of complementary HRM practices.

Note that, in general, adding agents with heterogeneous, nonstandard preferences to the problem will make the contracting environment become more complicated because all of these preferences are relatively difficult to measure, while firms' optimal strategies crucially depend on the composition of the workforce. Hence, HRM practices that decrease those information asymmetries and in turn increase matching quality are also a potential source of complementary HRM practices. To achieve this, firms can utilize various methods – they can screen workers by offering menus of contracts and let them self-select into occupations, or carefully test and screen applicants. Another method would be to just ask current

---

[13] This might not seem comparable to the aforementioned practices at first glance. However, gym membership and attendance have been in the focus of studying time-inconsistent behavior in the field. At the same time, gym memberships are a prime example of nonwage benefits that employers are offering with increasing frequency and that attract specific types of workers (see Lazear and Shaw, 2007). See Oyer (2008) for a discussion of factors that influence firms' benefit choices.

employees to suggest fitting applicants, or, in other words, request referrals. Recently, this mechanism has been receiving increased attention in several field experiments focusing either on different types of incentives for providing referrals (Beaman and Magruder, 2012), performance differences between referred and unreferred employees (Pallais and Sands, 2016), or other organizationally relevant outcomes such as turnover (Burks et al., 2015). Summing up, these studies emphasize that efficient matching of types matters and in turn may be a reason for the use of referrals. However, the aforementioned studies focus solely on referrals and therefore are silent about other HRM practices that firms employ. In contrast to that, our model predictions rely on the interactions between practices, and complementarities seem to be important for productivity outcomes. Consequentially, more research that explicitly refers to the complementarities between HRM practices and the role of screening in general and referral hiring in particular is needed.

4   Outlook

For answering the more general question how the potential explanatory power of HRM practices can best be studied, we suggest combining different methods and angles. Within this general agenda, WMS-style cross-sectional studies are important, as they map correlations and identify effective clusters of practices; to do so, in light of the above suggested role for behavioral insights, the scope of these cross-sectional studies has to be broadened to capture a wide set of practices employing a measurement model to assess potential complementarities. In parallel, insider-econometric-style studies help to provide in-depth insight into what matters in specific firms. In addition to these approaches, we argue that, in a first step, it is important to develop solid theoretical models yielding testable predictions, in particular including behavioral constructs. In a second step, these predictions are tested using both laboratory and field experiments to establish causality for theoretically suggested mechanisms.

One example for how we envision progress in research on organizations, management practices, and interactions is the Organizations Research Group (ORG),[14] an initiative at the University of Munich with the goal to explicitly study these topics by combining theory, lab and field experiments, and large-scale survey data. Hence, an important feature of ORG is the Organizational Observatory (O^2), a large-scale survey that aims at collecting rich primary data on organizations, their structure, and their design, with an emphasis on management practices that have not been studied extensively in the WMS, while striving for comparable data quality by utilizing the same survey methodology. Already started and being continuously improved, O^2 is ultimately planned to have a panel structure and to combine firm-level data on practices with employee-level data.[15] Beyond generating survey

---

[14] http://www.organizations-research-group.uni-muenchen.de/index.html.

[15] Related to this, we want to stress that there are some other groups of researchers that try to investigate the same kinds of questions by producing high-quality, detailed

data, ORG also has the aim to serve as hub for field experiments with and in firms. In combination with method-based training for graduate students through structured graduate programs like the International Graduate Program "Evidence-Based Economics,"[16] ORG provides expertise for interested firms and organizations to conduct randomized interventions. Thus, we hope for ORG to be an integral part of the integrated research on organizations that we envision in the future.

Summing up, by investigating the role of HRM practices in explaining persistent productivity differences and exploring the concept of complementarities between them, we have shown that integrating different research strands and employing different methods helps to gain new insights and to understand the field more thoroughly. Especially for the role of complementarities of HRM practices, we have demonstrated how linking seemingly unrelated results from behavioral-economics research on social preferences can be fruitful for further research in this area. However, by giving an overview of different issues and upcoming topics like determinants of productivity and referral hiring, we have also highlighted that more research, and especially more comprehensive firm data, are needed.

To conclude, we briefly sketch the virtuous feedback cycle we envision as evolving between differing research methodologies. Consider the following example: cross-sectional studies like the WMS, the LPP, or O² provide the possibility to relate within-firm patterns to market-level characteristics like market structure or the intensity of competition. So far, models of behavioral aspects have been largely silent on this topic, which is likely to have important effects on internal organization choices. That is why we hope that empirical findings from these surveys will stimulate theoretical research. Again, this theorizing will generate new predictions that have to be tested empirically. Here, experimental studies have the role of establishing hypothesized causal mechanisms, while insider-econometrics studies and new modules in surveys can again help to identify practices and complementarities among them.

Appendix

A.1   *Detailed Exposition of the Model*

Benchmark Case.   The principal faces the following optimization problem when implementing a_H:

$$\max_{w,b,\hat{a}} (1-\pi_H)(q_1-w) + \pi_H(q_2-w-b)$$

subject to

(IR) $(1-\pi_H)w + \pi_H(w+b) - c + \eta[(1-\hat{\pi})w + \hat{\pi}(w+b) - c(\hat{a})]ER(a_H) \geq 0,$

---

data; one example of this is the Linked Personnel Panel (LPP), which combines firm- and employee-level data (see Kampkötter and Sliwka, 2016, in this volume).

[16] http://evidence-based-economics.de/home.html.

$$(IC) \quad \begin{aligned} &(1-\pi_H)w + \pi_H(w+b) - c + \eta[(1-\hat{\pi})w + \hat{\pi}(w+b) - c(\hat{a})]ER(a_H) \\ &\geq (1-\pi_L)w + \pi_L(w+b) + \eta[(1-\hat{\pi})w + \hat{\pi}(w+b) - c(\hat{a})]ER(a_L), \end{aligned}$$

$$(EB) \quad \hat{a} = a_H.$$

The first and second constraints represent the standard individual rationality (IR) and incentive compatibility (IC) restrictions, respectively. The third constraint (EB) represents the restriction that beliefs have to be in equilibrium. More precisely, this means that the agent can reasonably expect the kindness of the proposed contract offer as his expected rent when choosing action $\hat{a}$. Thus, destructive and "babbling" equilibria are ruled out. However, note that it is sufficient to assume (EB) and solve in order to implement $\hat{a} = a_H$ (see the discussion in Englmaier and Leider, 2012a).

For deriving the optimal contract, it is helpful to first think about the properties of the constraints at the optimum. Considering (IR), it is straightforward to see that this constraint has to be binding: Suppose (IR) were slack.[17] Then, the principal could decrease w by the amount $\epsilon > 0$. For small enough ϵ, the principal would still manage to fulfill (IR) and at the same time increase her expected outcome without violating (IC), because (IC) depends on the bonus b (and thus on the wage spread between the states), which does not change. Thus, (IR) cannot be optimally slack.

Since (IR) is binding, the agent does not receive a rent, i.e., the reciprocal part of his utility is equal to zero and hence drops out of the problem. Thus, the optimal contract can easily be derived by solving the constraints of the standard problem for w and b.

Now, consider (IC). It can easily be seen that since the agent is risk-neutral and does not care about the wage spread, there exist an infinite number of optimal contracts implementing a_H at first-best cost for the principal, as long as $b \geq c/\Delta\pi$, where $\Delta\pi = \pi_H - \pi_L$. For reasons of simplicity and to make the first- and second-best solutions comparable, we focus on the case that the wages are set to solve (IC) with equality.

Limited Liability. Now, the optimization problem of the principal implementing a_H (when again assuming (EB) and solving to implement $\hat{a} = a_H$) takes the following form:

$$\max_{w,b} (1-\pi_H)(q_1 - w) + \pi_H(q_2 - w - b)$$

subject to

$$(IR) \quad (1-\pi_H)w + \pi_H(w+b) - c + \eta[(1-\pi_H)w + \pi_H(w+b) - c]ER(a_H) \geq 0,$$

$$(IC) \quad \begin{aligned} &(1-\pi_H)w + \pi_H(w+b) - c + \eta[(1-\pi_H)w + \pi_H(w+b) - c]ER(a_H) \\ &\geq (1-\pi_L)w + \pi_L(w+b) + \eta[(1-\pi_H)w + \pi_H(w+b) - c]ER(a_L), \end{aligned}$$

$$(LL) \quad w \geq \underline{w}.$$

---

[17] Note that, when we add limited liability as a source of inefficiency into the model below, (IR) may be optimally slack, as in Englmaier and Leider (2012a).

It is immediately clear that the principal is not able to implement the first-best contract in this case, because she would have to set $w < 0$, which would violate the limited-liability constraint (LL). Thus, we have to derive the second-best contracts, which we do by considering two cases – selfish and reciprocal agents.

Selfish Agent.   First, consider the optimal contract for a selfish agent ($\eta = 0$). Then, the optimization problem boils down to the standard case. Again considering the properties of the constraints in the optimum, it is easy to see that (LL) binds at the optimum. Additionally, (IC) is binding, which provides us with two equalities to solve for the optimal contract. The properties of this contract are summarized in the following proposition.

PROPOSITION 4 *Under hidden action with limited liability ($w \geq \underline{w} \geq 0$), the risk-neutral principal proposes the following contract to the risk-neutral, selfish agent ($\eta = 0$):*

$$w = \underline{w}, \quad b = \frac{c}{\Delta \pi}, \quad and \quad \hat{a} = a_H.$$

In doing so, she can induce the agent to exert a_H, but only at implementation cost

$$B^{SB}(a_H) = c + \underline{w} + \frac{\pi_L}{\Delta \pi} c.$$

The agent receives a limited-liability rent:

$$U(a_H) = \underline{w} + \frac{\pi_L}{\Delta \pi} c.$$

In this case, the first-best solution can no longer be employed. The constraint (LL) limits the range of possible wage payments. Comparing the bonuses in the first- and second-best cases reveals that they do not differ ($b = c/\Delta \pi$ in both cases), which makes sense when considering that (IC) is binding in both cases and thus the wage spread cannot be reduced further; however, the principal cannot charge negative wages, so she has to increase the base wage w and consequently pay a rent to the agent. In other words, while the principal is able to use both carrot and stick in the first-best case, she loses the opportunity of using the stick with limited liability and has to rely on the carrot alone to incentivize the agent.

Reciprocal Agent.   Now, consider a reciprocal agent with $\eta > 0$. We first illustrate this situation by a verbal argument: Since the agent is motivated not only by monetary incentives, but also by the reciprocal part of his utility that is triggered when he receives a rent exceeding his outside option, the principal has an additional instrument to induce a_H. Stated differently, the limited-liability rent has an additional incentive component and helps the principal to align the agent's interests to some degree with her own.

Indeed, solving the optimization problem in the same fashion as the standard limited-liability problem yields the same base wage w, but the bonus payment b, and thus also the wage differential between the states of the world, are reduced.[18]

### A.2 Additive Formulation of Worker Type and HRM Practices

Consider a simple, linear relationship between inherent and influenceable reciprocity: $\eta = \eta_1 + \eta_2$. Then, one can simply substitute η with $\eta_1 + \eta_2$ in Proposition 2 and again derive the comparative statics of the optimal bonus payment to see how an increase in η_2, for instance by investment in team building, influences b:

$$\frac{\partial b}{\partial \eta_2} = -\frac{\Delta \pi \Delta ER}{(\Delta \pi + (\eta_1 + \eta_2)\Delta ER \pi_H)^2}\left(\underline{w} + \frac{\pi_L}{\Delta \pi}c\right) < 0.$$

The second derivative and the cross derivative are both globally positive, indicating diminishing marginal returns to an increase in η:

$$\frac{\partial^2 b}{\partial \eta_2^2} = \frac{\partial^2 b}{\partial \eta_1 \eta_2} = \frac{\Delta \pi (\Delta ER)^2 \pi_H}{(\Delta \pi + (\eta_1 + \eta_2)\Delta ER \pi_H)^3}\left(\underline{w} + \frac{\pi_L}{\Delta \pi}c\right) > 0.$$

Figure A1

Dependence of Optimal Bonus on η in Additive Model
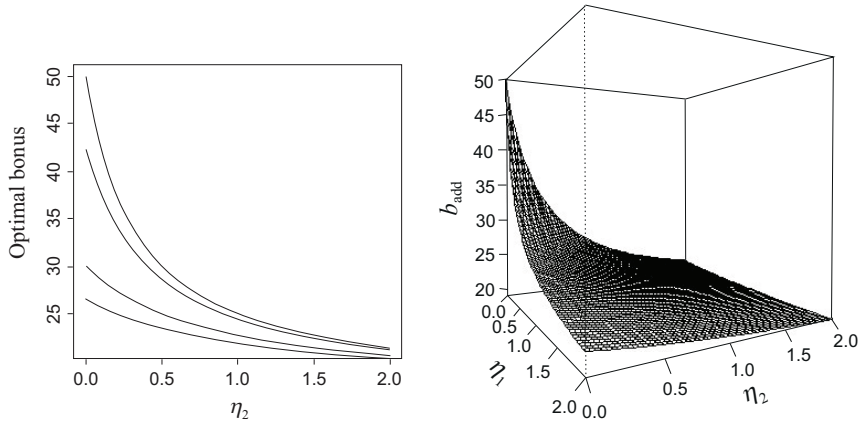
Notes: The figure depicts the optimal bonus as a function of η_1 and η_2, the parameters capturing the innate reciprocal inclination of a worker and the HRM policies targeted at increasing this reciprocal inclination towards the firm, for an additive formulation $\eta = \eta_1 + \eta_2$.

In this case, the two components of the reciprocity term act as substitutes, i.e., the firm can lower the bonus either by hiring workers with an inherent concern for

---

[18] Note that the bonus is decreasing in η. Assuming $c \geq \underline{w}$ is sufficient to ensure a positive b.

reciprocity in the first place, or by triggering reciprocal concerns later on through investments in HRM practices. This relationship can also be seen in the plot of the optimal bonus against η_1 and η_2 in Figure A1.

References

Abowd, J. M., J. Haltiwanger, R. Jarmin, J. Lane, P. Lengermann, et al. (2005), "The Relation among Human Capital, Productivity, and Market Value: Building up from Micro Evidence," in: C. Corrado, J. Haltiwanger, and D. Sichel (eds.), *Measuring Capital in the New Economy*, Studies in Income and Wealth, Volume 65, The University of Chicago Press, Chicago (IL) and London, pp. 153–204.

Akerlof, G. A., and R. E. Kranton (2000), "Economics and Identity," *The Quarterly Journal of Economics*, 115(3), 715–753.

— and — (2005), "Identity and the Economics of Organizations," *The Journal of Economic Perspectives*, 19(1), 9–32.

Ariely, D., E. Kamenica, and D. Prelec (2008), "Man's Search for Meaning: The Case of Legos," *Journal of Economic Behavior & Organization*, 67(3–4), 671–677.

Ashraf, N., O. Bandiera, and S. S. Lee (2014), "Awards Unbundled: Evidence from a Natural Field Experiment," *Journal of Economic Behavior & Organization*, 100, 44–63.

Athey, S., and S. Stern (2002), "The Impact of Information Technology on Emergency Health Care Outcomes," *The RAND Journal of Economics*, 33(3), 399–432.

Auriol, E., and R. Renault (2008), "Status and Incentives," *The RAND Journal of Economics*, 39(1), 305–326.

Autor, D. H., F. Levy, and R. J. Murnane (2002), "Upstairs, Downstairs: Computers and Skills on Two Floors of a Large Bank," *Industrial & Labor Relations Review*, 55(3), 432–447.

Azmat, G., and N. Iriberri (2010), "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students," *Journal of Public Economics*, 94(7–8), 435–452.

Bandiera, O., I. Barankay, and I. Rasul (2007), "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment," *The Quarterly Journal of Economics*, 122(2), 729–773.

—, —, and — (2009), "Social Connections and Incentives in the Workplace: Evidence from Personnel Data," *Econometrica*, 77(4), 1047–1094.

—, L. Guiso, A. Prat, and R. Sadun (2015), "Matching Firms, Managers and Incentives," *Journal of Labor Economics*, 33(3), 623–681.

—, V. Larcinese, and I. Rasul (2015), "Blissful Ignorance? A Natural Experiment on the Effect of Feedback on Students' Performance," *Labour Economics*, 34, 13–25.

Barankay, I. (2011a), "Gender Differences in Productivity Responses to Performance Rankings: Evidence from a Randomized Workplace Experiment," unpublished Manuscript, Wharton School of Business, University of Pennsylvania.

— (2011b), "Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment," unpublished Manuscript, Wharton School of Business, University of Pennsylvania.

Baron, J. N., and D. M. Kreps (1999), *Strategic Human Resources: Frameworks for General Managers*, John Wiley & Sons, New York.

Bartling, B., E. Fehr, and K. M. Schmidt (2012), "Screening, Competition, and Job Design: Economic Origins of Good Jobs," *The American Economic Review*, 102(2), 834–864.

Beaman, L., and J. Magruder (2012), "Who Gets the Job Referral? Evidence from a Social Networks Experiment," *The American Economic Review*, 102(7), 3574–3593.

Becker, B., and B. Gerhart (1996), "The Impact of Human Resource Management on Organizational Performance: Progress and Prospects," *The Academy of Management Journal*, 39(4), 779–801.

Becker, B. E., and M. A. Huselid (2006), "Strategic Human Resources Management: Where do we Go from Here?" *Journal of Management*, 32(6), 898–925.

Benkard, C. L. (2000), "Learning and Forgetting: The Dynamics of Aircraft Production," *The American Economic Review*, 90(4), 1034–1054.

Berg, N. A., and N. D. Fast (1975), "The Lincoln Electric Company," Case Study 376-028, Harvard Business School, Boston (MA).

Bernheim, B. D. (1994), "A Theory of Conformity," *Journal of Political Economy*, 102(5), 841–877.

Besley, T., and M. Ghatak (2005), "Competition and Incentives with Motivated Agents," *The American Economic Review*, 95(3), 616–636.

— and — (2008), "Status Incentives," *The American Economic Review*, 98(2), 206–211.

Blanes i Vidal, J., and M. Nossol (2011), "Tournaments without Prizes: Evidence from Personnel Records," *Management Science*, 57(10), 1721–1736.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013), "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 128(1), 1–51.

—, R. Lemos, R. Sadun, D. Scur, and J. Van Reenen (2014), "The New Empirical Economics of Management," *Journal of the European Economic Association*, 12(4), 835–876.

—, R. Sadun, and J. V. Reenen (2012), "Americans do IT Better: US Multinationals and the Productivity Miracle," *The American Economic Review*, 102(1), 167–201.

— and J. Van Reenen (2007), "Measuring and Explaining Management Practices across Firms and Countries," *The Quarterly Journal of Economics*, 122(4), 1351–1408.

Bowles, S. (1998), "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions," *Journal of Economic Literature*, 36(1), 75–111.

Bradler, C., R. Dur, S. Neckermann, and A. Non (2013), "Employee Recognition and Performance: A Field Experiment," Discussion Paper 13-017, Centre for European Economic Research (ZEW), Mannheim.

Brynjolfsson, E., and P. Milgrom (2013), "Complementarity in Organizations," in: Gibbons and Roberts (eds.) (2013, pp. 11–55).

Burks, S. V., B. Cowgill, M. Hoffman, and M. Housman (2015), "The Value of Hiring through Employee Referrals," *The Quarterly Journal of Economics*, 130(2), 805–839.

Cabrales, A., R. Miniaci, M. Piovesan, and G. Ponti (2010), "Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts," *The American Economic Review*, 100(5), 2261–2278.

Camerer, C. F., and R. A. Weber (2013), "Experimental Organizational Economics," in: Gibbons and Roberts (eds.) (2013, pp. 213–262).

Cappelli, P., and D. Neumark (2001), "Do 'High-Performance' Work Practices Improve Establishment-Level Outcomes?" *Industrial & Labor Relations Review*, 54(4), 737–775.

Cassar, L. (2014), "Optimal Contracting with Endogenous Project Mission," Working Paper 150, Department of Economics, University of Zurich.

Charness, G., D. Masclet, and M. C. Villeval (2014), "The Dark Side of Competition for Status," *Management Science*, 60(1), 38–55.

Chen, Y., F. M. Harper, J. Konstan, and S. X. Li (2010), "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens," *The American Economic Review*, 100(4), 1358–1398.

Cohn, A., E. Fehr, and L. Goette (2015), "Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment," *Management Science*, 61(8), 1777–1794.

Delery, J. E. (1998), "Issues of Fit in Strategic Human Resource Management: Implications for Research," *Human Resource Management Review*, 8(3), 289–309.

— and D. H. Doty (1996), "Modes of Theorizing in Strategic Human Resource Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions," *The Academy of Management Journal*, 39(4), 802–835.

Delfgaauw, J., and R. Dur (2007), "Signaling and Screening of Workers' Motivation," *Journal of Economic Behavior & Organization*, 62(4), 605–624.

DellaVigna, S. (2009), "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47(2), 315–372.

Dickinson, D., and M.-C. Villeval (2008), "Does Monitoring Decrease Work Effort? The Complementarity between Agency and Crowding-Out Theories," *Games and Economic Behavior*, 63(1), 56–76.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2009), "Homo Reciprocans: Survey Evidence on Behavioural Outcomes," *The Economic Journal*, 119(536), 592–612.

—, —, —, —, J. Schupp, and G. G. Wagner (2011), "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, 9(3), 522–550.

Ellingsen, T., and M. Johannesson (2007), "Paying Respect," *The Journal of Economic Perspectives*, 21(4), 135–150.

Englmaier, F., M. Fahn, and M. A. Schwarz (2015), "Long-Term Employment Relations when Agents Are Present Biased," Paper presented at the CESifo Area Conference on Behavioral Economics, October 23–24, https://www.cesifo-group.de/dms/ifodoc/docs/Akad_Conf/CFP_CONF/CFP_CONF_2015/be15-Schmidt/Paper/be15_Englmaier.pdf, accessed January 12, 2016.

—, T. Kolaska, and S. Leider (2015), "Reciprocity in Organizations: Evidence from the WERS," Working Paper 5168, Center for Economic Studies & ifo Institut (CESifo), Munich.

— and S. Leider (2012a), "Contractual and Organizational Structure with Reciprocal Agents," *American Economic Journal: Microeconomics*, 4(2), 146–183.

— and — (2012b), "Gift Exchange in the Lab – It is Not (only) How Much you Give ...," available at http://www-personal.umich.edu/~leider/Papers/GiftExchange_Lab.pdf, accessed on January 12, 2016.

— and — (2012c), "Managerial Payoff and Gift Exchange in the Field," Working Paper 3707, Center for Economic Studies & ifo Institut (CESifo), Munich.

—, S. Strasser, and J. Winter (2014), "Worker Characteristics and Wage Differentials: Evidence from a Gift-Exchange Experiment," *Journal of Economic Behavior & Organization*, 97, 185–203.

Falk, A., and M. Kosfeld (2006), "The Hidden Costs of Control," *The American Economic Review*, 96(5), 1611–1630.

Fehr, E., and K. M. Schmidt (2006), "The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories," in: S.-C. Kolm and J. M. Ythier (eds.), *Handbook of the Economics of Giving, Reciprocity and Altruism: Applications*, Volume 1, North-Holland, Amsterdam, pp. 615–691.

Gerhart, B. (2007), "Horizontal and Vertical Fit in Human Resource Systems," in: C. Ostroff and T. A. Judge (eds.), *Perspectives on Organizational Fit*, Lawrence Erlbaum Associates, Taylor & Francis Group, New York, pp. 317–348.

Gibbons, R., and J. Roberts (eds.) (2013), *The Handbook of Organizational Economics*, Princeton University Press, Princeton (NJ).

Grant, A. M. (2008), "The Significance of Task Significance: Job Performance Effects, Relational Mechanisms, and Boundary Conditions," *Journal of Applied Psychology*, 93(1), 108–124.

Gubler, T., I. Larkin, and L. Pierce (2013), "The Dirty Laundry of Employee Award Programs: Evidence from the Field," Working Paper 13-069, Harvard Business School, Boston (MA).

Hamilton, B. H., J. A. Nickerson, and H. Owan (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.

Holmstrom, B., and P. Milgrom (1994), "The Firm as an Incentive System," *The American Economic Review*, 84(4), 972–991.

Hoogveld, N., and N. Zubanov (2014), "The Power of (No) Recognition: Experimental Evidence from the University Classroom," Discussion Paper 7953, IZA, Bonn.

Hsieh, C.-T., and P. J. Klenow (2009), "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Huang, F., and P. Cappelli (2010), "Applicant Screening and Performance-Related Outcomes," *The American Economic Review*, Papers and Proceedings, 100(2), 214–218.

Huselid, M. A. (1995), "The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance," *The Academy of Management Journal*, 38(3), 635–672.

Ichniowski, C., and K. Shaw (2003), "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices," *The Journal of Economic Perspectives*, 17(1), 155–180.

— and — (2013), "Insider Econometrics," in: Gibbons and Roberts (eds.) (2013, pp. 263–314).

—, —, and G. Prennushi (1997), "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *The American Economic Review*, 87(3), 291–313.

Innes, R. D. (1990), "Limited Liability and Incentive Contracting with Ex-Ante Action Choices," *Journal of Economic Theory*, 52(1), 45–67.

Jackson, S. E., R. S. Schuler, and K. Jiang (2014), "An Aspirational Framework for Strategic Human Resource Management," *The Academy of Management Annals*, 8(1), 1–56.

Jiang, K., D. P. Lepak, K. Han, Y. Hong, A. Kim, and A.-L. Winkler (2012), "Clarifying the Construct of Human Resource Systems: Relating Human Resource Management to Employee Performance," *Human Resource Management Review*, 22(2), 73–85.

Jones, D., C. Willness, and S. Madey (2014), "Why are Job Seekers Attracted by Corporate Social Performance? Experimental and Field Tests of Three Signal-Based Mechanisms," *The Academy of Management Journal*, 57(2), 383–404.

Kampkötter, P., and D. Sliwka (2016), "On the Complementary Use of Experiments and Field Data to Evaluate Management Practices: The Case of Subjective Performance Evaluations," *Journal of Institutional and Theoretical Economics (JITE)*, 172(2), forthcoming.

Kepes, S., and J. E. Delery (2007), "HRM Systems and the Problem of Internal Fit," in: P. Boxall, J. Purcell, and P. Wright (eds.), *Oxford Handbook of Human Resource Management*, Oxford University Press, Oxford, pp. 385–404.

Kosfeld, M., and S. Neckermann (2011), "Getting More Work for Nothing? Symbolic Awards and Worker Performance," *American Economic Journal: Microeconomics*, 3(3), 86–99.

—, —, and X. Yang (2014), "Knowing that you Matter, Matters! The Interplay of Meaning, Monetary Incentives, and Worker Recognition," Discussion Paper 14-097, Centre for European Economic Research (ZEW), Mannheim.

Larkin, I., and S. Leider (2012), "Incentive Schemes, Sorting, and Behavioral Biases of Employees: Experimental Evidence," *American Economic Journal: Microeconomics*, 4(2), 184–214.

Lazear, E. P. (2000), "Performance Pay and Productivity," *The American Economic Review*, 90(5), 1346–1361.

— and K. L. Shaw (2007), "Personnel Economics: The Economist's View of Human Resources," *The Journal of Economic Perspectives*, 21(4), 91–114.

Leuven, E., H. Oosterbeek, R. Sloof, and C. van Klaveren (2005), "Worker Reciprocity and Employer Investment in Training," *Economica*, 72(285), 137–149.

Lockett, A., and S. Thompson (2001), "The Resource-Based View and Economics," *Journal of Management (JOM)*, 27(6), 723–754.

McAfee, A. (2002), "The Impact of Enterprise Information Technology Adoption on Operational Performance: An Empirical Investigation," *Production and Operations Management*, 11(1), 33–53.

Milgrom, P., and J. Roberts (1992), *Economics, Organization & Management*, Prentice-Hall, Englewood Cliffs (NJ).

— and — (1995), "Complementarities and Fit Strategy, Structure, and Organizational Change in Manufacturing," *Journal of Accounting and Economics*, 19(2), 179–208.

Mischel, W., and Y. Shoda (1995), "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure," *Psychological Review*, 102(2), 246–268.

O'Donoghue, T., and M. Rabin (1999a), "Doing it Now or Later," *The American Economic Review*, 89(1), 103–124.

— and — (1999b), "Incentives for Procrastinators," *The Quarterly Journal of Economics*, 114(3), 769–816.

Oyer, P. (2008), "Salary of Benefits?" *Research in Labor Economics*, 28, 429–467.

Pallais, A., and E. G. Sands (2016), "Why the Referential Treatment? Evidence from Field Experiments on Referrals," *Journal of Political Economy*, forthcoming.

Pfeffer, J. (1994), *Competitive Advantage through People: Unleashing the Power of the Work Force*, Harvard Business School Press, Boston (MA).

— (2007), "Human Resources from an Organizational Behavior Perspective: Some Paradoxes Explained," *The Journal of Economic Perspectives*, 21(4), 115–134.

Sautmann, A. (2013), "Contracts for Agents with Biased Beliefs: Some Theory and an Experiment," *American Economic Journal: Microeconomics*, 5(3), 124–156.

Schuler, R. S. (1989), "Strategic Human Resource Management and Industrial Relations," *Human Relations*, 42(2), 157–184.

— and S. E. Jackson (1987), "Linking Competitive Strategies with Human Resource Management Practices," *The Academy of Management Executive*, 1(3), 207–219.

Segal, C. (2012), "Working when No One is Watching: Motivation, Test Scores, and Economic Success," *Management Science*, 58(8), 1438–1457.

Shearer, B. (2004), "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *The Review of Economic Studies*, 71(2), 513–534.

Sliwka, D. (2007), "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes," *The American Economic Review*, 97(3), 999–1012.

Syverson, C. (2004), "Product Substitutability and Productivity Dispersion," *The Review of Economics and Statistics*, 86(2), 534–550.

— (2011), "What Determines Productivity?" *Journal of Economic Literature*, 49(2), 326–365.

Wolf, E., and T. Zwick (2008), "Reassessing the Productivity Impact of Employee Involvement and Financial Incentives," *Schmalenbach Business Review (sbr)*, 60(2), 160–181.

Woods, S. A., F. Lievens, F. De Fruyt, and B. Wille (2013), "Personality across Working Life: The Longitudinal and Reciprocal Influences of Personality on Work," *Journal of Organizational Behavior*, 34(S1), S7–S25.

Florian Englmaier   Katharina Schüßler
Department of Economics   Department of Economics
Kaulbachstr. 45   Kaulbachstr. 45
80539 Munich   80539 Munich
Germany   Germany
florian.englmaier@econ.lmu.de   katharina.schuessler@econ.lmu.de

Structural Estimation and Experiments: Applications to Contracting Models

by

Charles Bellemare, Steeve Marchand, and Bruce Shearer[*]

Structural estimation requires the specification of behavioral models that can be used to conduct ex ante policy evaluations and welfare analysis. Experiments generate data by exogenously varying key variables to measure outcomes under various treatment conditions. Gains from combining experiments and structural estimation can be considerable. We illustrate these gains through two recent papers (Bellemare and Shearer, 2011, 2013) in the area of contracts and compensation systems. In both papers, the combination of structural modeling and experiments is essential – experiments cannot be conducted to implement all possible treatments of interest, while structural estimation using naturally occurring (payroll) data either is infeasible or requires restrictive modeling assumptions. (JEL: J33, M52, C93)

1 Introduction

The economic analysis of compensation systems and contracts has evolved greatly since the 1970s. This is largely due to the development of theories of the firm incorporating information asymmetries. These models explicitly recognize that workers can control their productivity through their effort choices. The divergence of interests between firms and workers over effort levels leads to problems of moral hazard and adverse selection. Optimal compensation systems maximize firm profits, aligning interests by providing incentives for the worker to supply effort. Laffont and Martimort (2002), Bolton and Dewatripont (2004), and Salanié (2005)

---

[*] Charles Bellemare (corresponding author): Département d'économique, Université Laval, CIRPÉE, Canada, and CESiFo, Munich, and IZA, Bonn, Germany; Steeve Marchand: Département d'économique, Université Laval, CIRPÉE; Bruce Shearer: Département d'économique, Université Laval, CIRPÉE, CIRANO, Montréal (Québec), Canada, and IZA.

contain excellent summaries of this literature. The last two decades have seen a growing empirical literature that investigates how different compensation systems affect worker productivity and firm profits. Much early work exploited personnel data from the payroll records of firms. Such data allow researchers to observe worker productivity under different compensation systems; see, for example, Lazear (2000). More recently, many economists have turned to field experiments to conduct empirical work on compensation issues. Examples include Shearer (2004), Gneezy and List (2006), Fehr and Goette (2007), and Bandiera, Barankay, and Rasul (2007). Field experiments allow researchers to change the compensation system exogenously within a real-world environment and to measure the effect of such changes on worker productivity and behavior. Many proponents of experiments point to their advantages in attaining internal validity and in identifying causal effects; see, for example, Angrist and Pischke (2009).[1] We argue for a much broader interpretation of these advantages, one that also includes identifying the structural parameters of behavioral models, permitting ex ante evaluation of compensation systems and welfare analysis.

Structural models fully specify the environment in which individuals make decisions, including preferences, technology, constraints, and the sources of randomness. Reiss and Wolak (2007) and Keane, Todd, and Wolpin (2011) present discussions of the general approach. Within the context of contracting models, optimal worker behavior provides a mapping from structural parameters to worker output, conditional on the contract. This mapping forms the basis of identification. Structural parameters that are policy-invariant can be used to simulate the reaction of workers to other contracts, or alternative economic environments. Examples of research that uses the structural approach to investigate contracting models using payroll data include Ferrall and Shearer (1999), Paarsch and Shearer (1999), and Copeland and Monnet (2009).

Experiments add value to structural models. Standard economic theory suggests that the firm chooses compensation systems to maximize its profits, implying that any naturally occurring variation in contracts is endogenous. This point was made by Paarsch and Shearer (1999), who found piece rates and productivity to be negatively correlated in payroll data.[2] It complicates the use of payroll records to investigate compensation issues, as it requires the researcher to model explicitly the firm's choice of the compensation system; see, for example, Paarsch and Shearer (2000). The assumptions that are inherent in such an exercise form one of the major sources of criticism of the structural approach; see, for example, Angrist and Pischke (2010). Experiments provide exogenous variation in the compensation system and hence remove the need for this step, reducing (but not eliminating) the

---

[1] Much of the discussion over the benefits of experiments takes place within the context of the evaluation of social programs; see, for example, Ashenfelter (1987); Burtless (1995).

[2] In their particular example, this was due to the firm setting the piece rate as a function of elements that affected worker productivity but were unobservable to the econometrician.

modeling assumptions necessary to conduct structural analysis.[3] At the same time, reducing modeling assumptions can significantly lessen the computational burden imposed by structural estimation. This burden is often seen as an important barrier to entry in the area of structural estimation (see Keane, 2010, for a discussion). Finally, experiments provide a natural way to generate holdout samples that can be used to assess the external validity of structural models. This requires estimating a model holding out a subset of the data (control data, treatment data, or a mix of both) and using the model estimates to predict choices in the holdout sample. Recent work validating structural models with holdout samples generated using experiments includes Todd and Wolpin (2006) and Bellemare, Kröger, and van Soest (2008).

Structural estimation also adds value to experimental studies on compensation, both in generalizing experimental results and in conducting welfare analysis. Field experiments typically offer the researcher a limited amount of variation in compensation policies within a specific setting. Experiments are expensive, and the contractual variation that is possible is often restricted ex ante. For example, firms may be hesitant to cut wages, fearing adverse effects on worker morale. Alternatively, the experiment may be conducted within a given labor-market setting, which is beyond the researcher's control. This raises questions about the effects of alternative compensation policies or alternative environments on worker productivity and firm profits. Given estimates of the structural parameters of a particular model, worker behavior and profits can be simulated under alternative contracts or environments by changing the contract (or environment) and solving for optimal behavior under the new incentive structure. Examples of empirical work that combines structural modeling with experimental data to investigate contractual performance include Shearer (2004), Paarsch and Shearer (2009), and Bellemare and Shearer (2011, 2013). Structural modeling has been combined with experimental data in other contexts, in particular to evaluate and predict social policies; see, for example, Todd and Wolpin (2006); Attanasio, Meghir, and Santiago (2012); Ferrall (2012).[4]

We illustrate the complementarities between structural models and field experiments in evaluating contracts and compensation with two examples drawn from the research of Bellemare and Shearer. Both examples demonstrate the ability to generalize experimental results to alternative labor-market settings and to predict the profitability of contracts that are not observed within a given experiment. Our first example, Bellemare and Shearer (2011), considers gift exchange within the firm (Akerlof, 1982). In these models, firms benefit from the reciprocal nature of workers by offering wages above market-clearing levels. These models offer a way to rationalize persistent unemployment and downward price rigidities. The main

---

[3] It is worthwhile to compare Paarsch and Shearer (2000) with Shearer (2004) in this respect.

[4] Structural models can also be applied to data collected from laboratory experiments to identify preference parameters. See, for example, Bajari and Hortaçsu (2005); Bellemare, Kröger, and van Soest (2008); Brown, Flinn, and Schotter (2011).

empirical challenge in testing these models is to identify the part of wages that represents gifts. Experiments are especially useful in this context to clearly identify when gifts are given, their magnitudes, and how workers respond to those gifts. Recent field experiments have shown that while workers do respond positively to monetary gifts from their employer, increasing their productivity, the gift is not profitable (Gneezy and List, 2006; Bellemare and Shearer, 2009). This raises the issue whether or not alternative gifts, or gifts given in different circumstances (such as different labor markets), would be profitable. Bellemare and Shearer (2011) develop and estimate a structural model of reciprocity within the firm that permits them to provide answers to such questions. A key element of their model is the definition of the gift, specified as an increase in the expected earnings beyond market-clearing levels, holding effort fixed. The response of the worker to a gift depends on his/her reciprocity parameter as well as the effort levels that are induced in the absence of gifts. These effort levels are determined endogenously. High effort levels crowd out reciprocity because the marginal cost of additional effort is high. Bellemare and Shearer (2011) use data from field experiments conducted within a firm to estimate reciprocity and cost-of-effort parameters. They then use these parameters to predict how effort levels would change under different gift contracts and in different labor markets.

Our second example, Bellemare and Shearer (2013), looks at the potential benefits of matching workers to risk environments within a firm that pays its workers piece rates. The analysis is ex ante, since the firm does not match workers to conditions under its current personnel policy. The more risk-averse workers are, the higher are the costs of exposing them to risky environments.[5] If workers have heterogeneous risk preferences, then the firm could gain by matching the more risk-averse workers to less risky settings, while leaving workers who are less risk-averse (or are risk-loving) on the risky contracts. Bellemare and Shearer (2013) develop and estimate a structural model of contracting within a firm that takes account of risk preferences and risk levels. They solve their model for the supply functions of workers with respect to different risk environments. These functions define piece rates that ensure worker participation in a given contract as a function of his/her risk preferences and the level of risk. Bellemare and Shearer (2013) use a series of field experiments, conducted within the firm to identify these parameters and the supply functions. They then solve the matching problem by computing the profit-maximizing set of piece rates that allocates workers to different risk settings within the firm. Their results suggest that matching would generate a small increase in profits, given the planting conditions observed in the firm.

---

[5] This leads profit-maximizing firms to reduce incentive pay in such environments, leading to the risk–incentive trade-off of textbook models. An inability to detect such a trade-off in empirical work has puzzled economists (Prendergast, 2002). Yet, if workers are heterogeneous in risk preferences, such a trade-off does not necessarily occur (Ackerberg and Botticini, 2002). Contracts with high levels of risk attract individuals who are risk-neutral or risk-loving, requiring no trade-off with incentives.

The rest of the paper is organized as follows. Section 2 presents some institutional details of the industry in which our contracting studies take place. In section 3 we present the basic model that we use to study contracting issues. In section 4, we present our applications. Section 5 concludes.

2   Tree Planting

The experiments discussed in this paper were all conducted in a tree-planting firm located in British Columbia, Canada. In this industry, several firms compete during the autumn for planting rights on various recently logged blocks of land across the province. Firms are normally able to inspect blocks before submitting their bids to the government. The firm with the winning bid typically schedules planting for the following planting season, which normally takes place during spring and summer. At this time, workers are hired to plant seedlings on the blocks of land. They move around on foot, carrying the seedlings in a sack that fits around their waist.

The workers in this firm are paid piece rates, based on their daily productivity (the number of trees planted). The contract specifies a price per tree planted that is paid to each worker. There is no base wage. Productivity depends on the effort of the worker and the planting conditions found on a particular block. Terrain that is steep or contains compact or rocky soil is more difficult to plant, slowing the planters down. These planting conditions can vary a great deal from block to block. For each block to be planted, the firm decides on a piece rate to be paid to the planters. Blocks that are more difficult to plant require higher piece rates to attract workers. The piece rate applies to all planting done on a block; no systematic matching of workers to planting conditions occurs within the firm.

Conditions vary within blocks as well. For example, some parts of a given block may be characterized by rocky soil under the surface, making planting more difficult. Given that the firm cannot know completely the planting conditions for the whole block, and given that the contract is constant within each block, some planters will invariably end up working in more difficult conditions under the same contract. These random elements expose planters to daily income risk.

The compensation policy of the firm has changed very little over the years, with no systematic exploration of alternative forms of contracts. Yet, several firms compete in this market to attract and retain productive workers, and to complete contracts quickly and cost-effectively. Alternative contracts may help firms in this market reach these objectives. For example, the literature on gift exchange, for example, argues that firms can benefit by offering gifts in the form of wages above market-clearing levels, with reciprocal workers responding to gifts by raising their productivity. Firms may also benefit by letting workers sort themselves across blocks, with risk-averse workers accepting lower piece rates to work on blocks with lower daily income variability. It is not clear a priori that these alternative contracts can be profitable for firms in this market with a heterogeneous workforce characterized by varying levels of reciprocity, risk preferences, and ability. As we

argue in the following sections, the analysis of the performance of alternative contracts not yet implemented by the firm requires combining structural estimation methods with experiments.

3   Model Preliminaries

The two applications discussed below each specify a structural model of worker behavior in the firm. The two models have elements in common. In particular, we assume that the daily productivity of worker i on block j is determined by worker effort, E_{ij}, and a productivity shock, S_{ij}:

$$(1) \qquad\qquad Y_{ij} = E_{ij} S_{ij}.$$

The productivity shock represents planting conditions such as the hardness and the steepness of the ground. We assume that $\ln(S_{ij})$ follows a normal distribution with mean μ_j and variance σ_j^2. Throughout we assume that workers and the firm know the distribution of planting conditions. Workers accept or reject the offered contract before knowing the actual planting conditions to which they will be assigned, S_{ij}.

Contrary to the firm, workers are able to observe the planting conditions (a realization of S_{ij}) at the beginning of work on the block and can thus condition their effort on this information. This introduces asymmetric information between the workers and the firm, given that the latter cannot verify the actual work conditions of each worker.

Daily earnings are given by $W_{ij} = r_j Y_{ij}$, where r_j denotes the piece rate paid to all workers planting on block j. Workers are assumed to have CRRA utility functions and possibly reciprocal preferences (see, e.g., Rabin, 1993), the latter of which will be discussed in detail below. These preferences are defined over earnings on block j (W_{ij}) and effort (E_{ij}):

$$(2) \qquad\qquad U_i = \frac{1}{\delta_i}[W_{ij} - C_i(E_{ij}) + \beta_i G_{ij} \Pi_{ij}]^{\delta_i},$$

where

$$C_i(E_{ij}) = \frac{\kappa_i}{\eta} E_{ij}^{\eta}$$

denotes individual i's cost of effort, G_{ij} denotes the gift given by the firm to the worker, and Π_{ij} captures the kindness of the worker to the firm in response to the gift. Specific functional forms for G_{ij} and Π_{ij} are presented in section 4.1 below. The parameter β_i captures the strength of reciprocal preferences for worker i. With $\beta_i > 0$, a worker receiving a positive (negative) gift G_{ij} gets more utility if he/she increases (decreases) his/her effort relative to his/her effort level in the absence of gifts. Here, κ_i allows for individual-specific ability in tree planting and η measures the curvature of the cost function. The parameter δ_i denotes the risk-preference parameter of worker i.

Conditional on s_{ij}, a realization of S_{ij}, the worker selects effort to maximize utility. As we discuss below, the optimal effort for both applications is given by the following expression:

$$(3) \qquad \hat{e}_{ij} = \left[\frac{(r_j + \beta_i G_{ij})s_{ij}}{\kappa_i} \right]^{\gamma},$$

where $\gamma \equiv 1/(\eta - 1)$. The optimal effort is independent of risk preferences, given our assumption that workers observe conditions before selecting their effort level. Note, however, that the equilibrium effort will still be affected by risk through the participation constraint and the determination of r_j. In particular, we assume the firm sets r_j to satisfy the participation constraint of the marginal worker h, defined to be that worker who is indifferent between working and staying home.

4 Applications

4.1 Application 1: Profitability of Gift-Giving

Reciprocity models were first developed by anthropologists and sociologists (see Mauss, 1923; Homans, 1954) and were later introduced into economics by Akerlof (1982). These models are based on the principle of reciprocity, stating that individuals who receive gifts feel obliged to reciprocate with a return gift. Akerlof argued that workers will interpret wage increases above market-clearing levels as gifts from the firm. They will, in consequence, reciprocate by increasing their effort levels. Although Akerlof's model generated interest among economists, little empirical work immediately followed. This was, in large part, due to the difficulties involved in measuring the variables of interest. The most fundamental prediction associated with Akerlof's model concerns the principle of reciprocity, stating that workers will respond to a gift by increasing their productivity. Testing this prediction requires measuring the size of gifts (or who receives a gift) and how they react to it. Yet, measuring gifts is problematic in most data sets. Higher-than-average wages may reflect a gift or simply unobserved heterogeneity that affects productivity. Moreover, most data sets only contain information on wages, making it difficult to measure the effect of a gift on worker productivity.

Experiments solve the measurement problems by allowing researchers to generate gifts under controlled circumstances and to observe participants' reaction to those gifts, measuring the reciprocal responses as a result. For these reasons the majority of empirical work on gift exchange has been experimental. Most early studies were conducted in the laboratory. These studies typically generate a very robust relationship between gifts and worker effort (see for example Fehr, Kirchsteiger, and Riedl, 1993).

Questions over the generalizability of laboratory results and their relevance for real-world labor markets led some authors to conduct field experiments investigating reciprocity, e.g., Gneezy and List (2006); Falk (2007); Bellemare and Shearer

(2009). While a majority of studies find that workers do increase effort and productivity in response to an experimental gift, these gifts are seldom profitable to the principal. For example, Bellemare and Shearer (2009) found that a gift of $80 provided to workers in a tree-planting firm generated increased output worth only $17 to the firm. The lack of profitability of gift-giving raises questions about the usefulness of these models in explaining observed wages in real-world settings. However, these studies only consider the profitability of specific gifts, implemented within a specific experiment. The lack of profitability of these experimental gifts does not rule out the possibility that alternative gifts, or gifts given under alternative labor-market conditions, might be profitable. Providing answers to such questions requires predicting the manner in which workers will react to gifts in these alternative circumstances. Estimating structural behavioral models is a means by which economists can investigate such questions. These methods estimate (policy-invariant) parameters that generate worker responses to gifts, allowing the ex ante prediction of results outside the experimental environment.

4.1.1   The Structure of Gifts and Reciprocity

Bellemare and Shearer (2011) apply a variant of the model developed above to investigate gift exchange and reciprocity within firms when workers are risk-neutral ($\delta = 1$).[6] They define a gift as the gain in expected utility, holding effort fixed at pre-gift levels. Their definition is general and can accommodate gifts of different forms. For example, gifts can be generated through increases in fixed wages B_{ij} beyond market-clearing levels $\bar{\omega}$, but also through increases in the piece rate R_{ij} beyond market-clearing levels $\bar{r}_j$.

They show that gifts can be expressed in this setting as

$$(4) \qquad G_{ij}(R_{ij}, B_{ij}, \bar{r}_j) = B_{ij} + (R_{ij} - \bar{r}_j)\bar{r}_j^{\gamma} A_{ij},$$

where

$$A_{ij} = \frac{\mathrm{E}(S_{ij}^{\gamma+1})}{\kappa_i^{\gamma}}$$

represents the expectation of a productivity shock.

Note that the base wage that would be offered in the absence of gifts is set to zero in order to reproduce the particular setting studied by Bellemare and Shearer (2011). Note also that the gift simplifies to B_j if it is offered as a common base wage to all workers. From (2) it follows that reciprocal workers who are offered a gift reciprocate and return value Π_{ij} to the firm. We measure Π_{ij} as the increase in a worker's output relative to the output the worker would produce in the absence of gifts. This implies that $\Pi_{ij} = Y_{ij} - Y_{ij}^{ng}$, where Y_{ij} and Y_{ij}^{ng} denote output with and

---

[6]   Implications of relaxing this assumption to allow for risk-averse and risk-tolerant workers are discussed in Bellemare and Shearer (2011, section 9.1).

without the gift, respectively – the latter being evaluated at optimal effort measured in the absence of gifts. It follows that optimal effort is given by (3).

The structural model is composed of the equations (4), defining the gift, and (3), defining the optimal effort for a given gift, together with the production function (1) mapping from unobservable shocks and effort to observable output. A participation constraint ensures that workers are willing to work and provides a link between labor-market conditions and market-clearing compensation, $\bar{r}_j$. These equations and the firm's profits are functions of the utility-function parameters $\beta_i, \kappa_i, \gamma$; distributional parameters A_{ij}; and policy (or contractual) variables $R_{ij}, B_{ij}, \bar{r}_j$.

The structural model makes clear why the results of any given experiment may not generalize beyond the specific environment in which the experiment was conducted. First, the type (or size) of the gift may differ across experiments, which will change the effort decision of reciprocal workers. Second, and perhaps more subtly, the reciprocal response of the workers is likely to depend on the level of effort that is provided in the absence of gifts. In the particular setting studied by Bellemare and Shearer (2011), the state of the labor market affects effort through $\bar{r}$, the piece rate that must be paid in a competitive labor market. If the cost of effort is increasing and convex ($\gamma < 1$), then high effort levels in the absence of gifts raise the marginal cost of reciprocity. The reciprocal response to gifts will be crowded out by effort in the absence of gifts.

The utility-function parameters are policy-invariant – they do not change their values under different gifts. This provides the predictive force of the structural approach. Different types of gifts (or gifts in different labor markets) will generate effort levels (and output) that are functions of the *same* structural parameters as those generating responses within the experiment. Once these parameters have been estimated using experimental data, the effort, productivity, and profits can be predicted in other settings by solving the model for the optimal effort level under a different gift or different labor-market conditions. An experiment, varying a limited number of gifts, can be used to estimate the structural parameters, which then serve to predict the effort, productivity, and profits for other gifts.

4.1.2   Identification

Bellemare and Shearer (2011) show that computing expected profits for a given gift requires estimates of γ, the curvature of the cost of effort; β_i, the reciprocity parameter; and A_{ij}, the expected productivity shock. These parameters determine the reaction to incentives in the presence and in the absence of gifts.

The parameter γ captures the curvature of the cost-of-effort function. It determines the reaction of workers to incentives in the absence of gifts. Bellemare and Shearer (2011) identify γ using experimental data collected in the firm and first analyzed in Paarsch and Shearer (2009). In this experiment, several homogeneous treatment blocks were separated into two subplots. The firm paid the regular piece rate on one of the subplots, and a higher piece rate on the other subplot. This design ensured that the assigned piece rate r_j was independent of the planting con-

ditions s_{ij} of each block. What is more, each subplot was presented to workers as a separate block requiring a separate piece rate given the different planting conditions. As a result, the higher piece rates paid on the selected subplots of each block were not framed as a gift of the firm to its workers. Setting $G_{ij} = 0$ in (3) and substituting into the production function yields an expression for optimal productivity as a function of the piece rate and γ. Since the experiment varies the piece rate, holding planting conditions fixed, γ can be estimated through a direct comparison of productivity under different experimental piece rates. The estimated value of γ is 0.39 and is statistically significant at all levels of confidence. This estimate suggests that the cost-of-effort function $C_i(E_{ij})$ is convex.

Identification of the reciprocity parameter β_i exploits data analyzed from an experiment designed to measure worker response to gift-giving within the firm. These data were first analyzed in Bellemare and Shearer (2009). The experimental design had the firm present workers with a windfall gift of $80 on a selected morning before planting started. Importantly, workers were informed that this was a one-time gift that would not be repeated in the future. The gift was thus not framed as a signal that the firm adopted a new compensation policy. In the notation presented above, this implies $G_{ij}(R_{ij}, B_{ij}, \bar{r}_j) = B_j = 80$. Estimation of β_i was performed using nonlinear least squares. They first assume that β_i is constant across all planters and find an estimate $\hat{\beta}$ of 0.00071, which is statistically significant at a 5 % level. When allowing for heterogeneous reciprocity parameters $\hat{\beta}_i$, they find that reciprocity effects increase with work experience within the firm and decrease with increasing age. Moreover, the estimated value of β_i is positive and significantly different from zero for slightly more than half of the workers.

These experiments clearly demonstrate the advantage of using experimental methods to estimate structural parameters. For example, nonexperimental estimates of γ using payroll records are based on naturally occurring variation in the piece rate across contracts. Yet, such variation is likely to be endogenous, complicating identification; see, for example, Paarsch and Shearer (1999). An experiment varies the piece rate exogenously, giving rise to simple estimators that directly (and convincingly) identify γ. Similarly, estimating β_i is greatly facilitated through the use of experimental methods. Not only is the variation in the gift exogenous, but the experimental setting clearly identifies when a gift is presented.

4.1.3   Results

Bellemare and Shearer (2011) consider three types of gifts:

(i)   base-wage gifts, wherein the piece rate is fixed at the market-clearing rate $\bar{r}_j$, and base wages can be positive ($B_j > 0$);

(ii)   piece-rate gifts, wherein the piece rate is fixed above the market-clearing rate $\bar{r}_j$, and the base wage is zero;

(iii) composite gifts, wherein the piece rate is fixed above the market-clearing rate $\bar{r}_j$, and the base wage is positive.

Figure 1

Optimal Piece-Rate Gift (left) and Percent Increase in Profits (right)
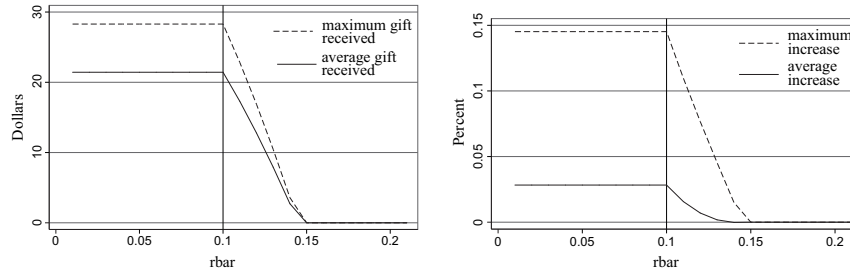
In all three cases they calculate the optimal gift that maximizes expected profits subject to the participation constraint of each worker. They investigate how the labor market affects the return to gift exchange by varying the market-clearing piece rate $\bar{r}_j$. They find that base-wage gifts would never be profitable, whatever the value of $\bar{r}_j$. This suggests that base-wage gifts have little importance in this firm.

Piece-rate gifts can be profitable if labor-market conditions are slack. This result is illustrated in Figure 1, presenting the optimal gift value and the resulting percentage increase in profit.

We see that gifts are not profitable when $\bar{r}_j > 0.15$ (the regular piece rate in the gift-giving experiment was 0.2) – the optimal gift value is zero, so profits are not affected. As $\bar{r}_j$ decreases toward 0.1, labor-market conditions become increasingly slack and gift-giving using piece rates becomes profitable. The diminishing return of G_{ij} with $\bar{r}_j$ is linked to the increasing marginal cost of effort. When $\bar{r}_j$ is low, incentives in the absence of gifts are low-powered and the employees exert low effort. In these circumstances, it is not costly to increase effort, and workers react strongly to gifts from their employer. On the contrary, when $\bar{r}_j$ is high, incentives are high-powered in the absence of gifts and employees exert high effort. Providing additional effort in response to a gift is very costly in this case, and workers react little to gifts from their employer.

The profitability of gift-giving levels off at values of $\bar{r}_j$ lower than 0.1. In our setting, $\bar{r}_j$ corresponds to the profit-maximizing regular piece rate in the absence of gifts if participation constraints are not binding. As a result, it is never optimal for the firm to set a regular piece rate below 0.1 in the absence of gifts. Figure 1 also reveals considerable heterogeneity in the profitability of piece-rate gifts across workers. In particular, while the average profit increase reaches at most 3 %, the maximum profit increase reaches almost 15 %. This heterogeneity follows from the fact that the same piece-rate gift will generate different gift values G_{ij} for workers because of their heterogeneous skills, which are incorporated in A_{ij}. It will also generate different responses due to heterogeneity in the reciprocal preferences β_i.

Figure 2

Optimal Composite Gift (left) and Percent Increase in Profits (right)
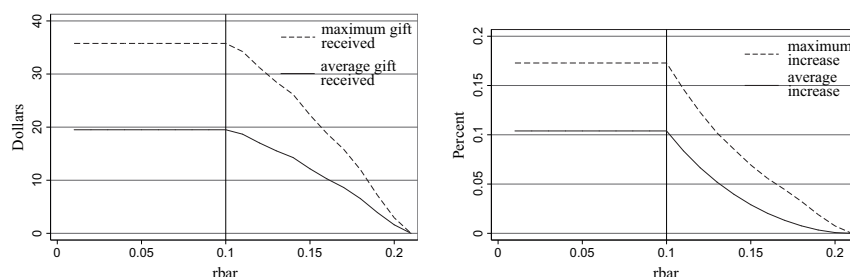
Source: Bellemare and Shearer (2011).

Composite gifts are a combination of base wages and piece-rate gifts. As in the case of piece-rate gifts, composite gifts are most effective in slack labor markets. Figure 2 presents the optimal composite gifts and the resulting profit increases.

The average optimal composite gift is near zero when $\bar{r}_j = 0.2$, which is the observed regular piece rate on experimental blocks. This suggests composite gifts would have a very small economic role in the firm's actual setting. As with the optimal piece-rate gift, however, the optimal gift increases when $\bar{r}_j$ decreases, reaching a ceiling when $\bar{r}_j$ drops to 0.1. Overall, optimal composite gifts lead to higher average profits than piece-rate gifts. In slack labor markets, where the competitive piece rate would be very low (and thus employees would exert low effort), the composite gift could increase the average profit per worker by up to 10 %. Given employees' heterogeneous responses to gifts, the increase in profit reaches around 17 % for some planters.

Interestingly, the optimal base-wage component of the gift is negative. In contrast, the piece-rate component is positive and sufficiently high to make the total gift value positive for each worker. The negative base-wage component is perhaps surprising from the point of view of gift exchange, but not from the point of view of contract theory. It reflects the ability of the firm to benefit from the incentives inherent in higher piece-rate gifts. These piece rates would lead to workers earning rents. The negative base wage allows the firm to recapture some of those rents.

The marginal benefit from applying structural econometric modeling to the gift experiment is clearly demonstrated on comparing these results with the reduced-form econometric approach followed by Bellemare and Shearer (2009). In that paper (as is typical of reduced-form approaches) the main concern was the estimation of the treatment effect: the effect of the gift on worker productivity and firm profits. Yet the conclusions of such an exercise are limited to the particular experimental environment and treatment. Bellemare and Shearer (2009) conclude that while the gift did have an effect on worker productivity, the gift was not profitable in the observed setting. Structural analysis permits generalization to other settings, predicting when gifts would be profitable and how such gifts should be formed.

4.2 Application 2: Cost of Risk and Benefits of Matching

The firm normally sets the piece rate on each block primarily on the basis of planting conditions, and subsequently assigns workers randomly across blocks available on a given day. A natural consequence of random assignment of workers to planting conditions is that the firm does not benefit from the possibility that risk-averse workers may be willing to pay (accept a lower piece rate) if they are able to sort themselves across blocks with varying levels of daily income risk. Recent evidence suggests that workers partly do in fact sort themselves across work environments based on their risk preferences (Ackerberg and Botticini, 2002; Bellemare and Shearer, 2010). As a result, sorting potentially reduces the cost of risk in the market if risk-tolerant workers are attracted to risky settings. Risk can nevertheless remain an important source of contracting costs when worker heterogeneity is extended to multiple dimensions and sorting is based on factors other than risk preferences. The cost of risk will then depend on the risk preferences of workers and the risk levels to which they are exposed.

Bellemare and Shearer (2013) combine structural estimation, payroll data, and experimental data to measure the cost of risk and the benefits of matching heterogeneous workers to risk levels. In contrast to the previous application, they consider piece-rate contracts only. However, workers are allowed to sort themselves across the different blocks available on a given day. Under matching, the firm sets piece rates by exploiting information about planting conditions, worker ability, and the distribution of risk preferences. In contrast, piece-rate setting with random assignment of workers to blocks does not require knowledge of the distribution of risk preferences – only the preferences of the marginal worker are necessary, as will be discussed below.

This study provides an example of how experimental methods can generate data that can be used as input to perform counterfactual predictions of the effect of sorting on matching. The firm in question did not match workers to working conditions. The evaluation of the benefits to matching is therefore ex ante, predicting the benefits of a personnel policy that has not yet been put into place. Reciprocity is not introduced in this paper, so we set $G_{ij} = 0$ in (2).

4.2.1 Cost of Risk to Workers

In the presence of heterogeneous preferences, earnings adjust to compensate the marginal worker for his/her differences in expected utility across contracts. Importantly, effort costs and risk both change across contracts. This implies that the observed earnings differential compensates for both changes in effort costs and risk. Calculating the cost of risk in this setting therefore requires measuring the earnings that are required to compensate for risk, holding effort constant.

To measure the cost of risk to workers on a given contract, we calculate the amount each worker is prepared to pay to eliminate risk on that contract, holding expected effort costs constant at optimal levels. We define $\overline{W}_{ij}$ to be worker i's certainty equivalent income on block j. Then $\overline{W}_{ij}$ provides the worker with the same

level of expected utility as he/she gains from working on block j under uncertainty, holding expected effort costs constant at the level implied by optimal behavior. The cost of risk, cr_{ij}, for individual i on block j is therefore obtained by subtracting $\overline{W}_{ij}$ from the equilibrium expected earnings predicted by our model, giving

$$(5) \qquad cr_{ij} = \bar{\omega}\left(\frac{\kappa_h}{\kappa_i}\right)^\gamma \left[\exp^{0.5(\gamma+1)^2\sigma_j^2(1-\delta_h)} - \exp^{0.5(\gamma+1)^2\sigma_j^2(\delta_i-\delta_h)}\right].$$

Inspection of (5) reveals the following: First, as expected, the cost of risk is zero in the absence of risk ($\sigma_j^2 = 0$). Second, the cost of risk is increasing (decreasing) in σ_j^2 if individual i is risk-averse (-loving). Third, the cost of risk is proportional to planting ability, given risk preferences. This is due to the fact that the moments of the earnings distribution depend on ability (relative to the marginal worker).

4.2.2 Benefits of Matching

The heterogeneity in risk preferences suggests that there are potential gains to the firm from matching workers to contracts based on risk conditions and preferences. These benefits can be evaluated from the estimates of the structural model. To fix ideas, consider two blocks L and H, both with the same average planting conditions but different variances such that $\sigma_H^2 > \sigma_L^2$. Suppose further that the firm would pay $\hat{r}_L$ and $\hat{r}_H$ on these blocks when randomly assigning workers to each block. The firm could instead profit by paying $\tilde{r}_l < \hat{r}_L$ on block L and allow workers to choose the block on which they want to plant, while still increasing their utility vis-à-vis block H. Whether or not actual gains are realized will depend on the change in behavior of the workers who self-select onto block L. As the piece-rate changes, their effort levels will change affecting firm profits. We now turn to calculating the effect on profits. Let the set $\Delta(\tilde{r}_L)$ contain all workers who will choose to plant on block L as a function of $\tilde{r}_L$:

$$\Delta(\tilde{r}_L) = \{\delta_i : \delta_i < \delta^*(\tilde{r}_L; \gamma, \delta_h, \sigma_H^2, \sigma_L^2)\}.$$

The threshold value $\delta^*(\tilde{r}_L; \gamma, \delta_h, \sigma_H^2, \sigma_L^2)$ depends on parameters that are all identified using either payroll data or the experimental data. It determines the supply of workers to block L as a function of the piece rate $\tilde{r}_L$.

To calculate the gains from matching, we consider the firm's expected daily profits that result from allowing workers to sort across a given high-variance block H and a low-variance block L in the firm. We denote these profits by $\pi_{t,H,L}^m$. We then calculate the profits from randomly allocating workers across these two blocks, denoted by $\pi_{t,H,L}^{nm}$. We denote the expected profit increase to matching workers between these two blocks by $\pi_{t,H,L}$.

To calculate the profits from not matching workers, we randomly allocate workers to block H and block L. Profits are then given by

$$(6) \qquad \pi_{t,H,L}^{nm} = \sum_{i\in\mathcal{H}}\pi_H^{nm} + \sum_{i\in\mathcal{L}}\pi_L^{nm},$$

where $\mathcal{H}$ denotes the set of workers who are randomly allocated to plant on block H, and $\mathcal{L}$ denotes the set of workers who are randomly allocated to plant on block L.

Under matching, the firm chooses $\tilde{r}_L$ to maximize

$$(7) \qquad \pi_{t,H,L}^{m} = \sum_{i \in \Delta(\tilde{r}_L)} \pi_H^m + \sum_{i \notin \Delta(\tilde{r}_L)} \pi_L^m.$$

The increase in expected profit from matching between block j and L is then given by

$$(8) \qquad \pi_{t,H,L} = \frac{\pi_{t,H,L}^{m} - \pi_{t,H,L}^{nm}}{\pi_{t,H,L}^{nm}}.$$

The expected profits with and without matching depend on the distribution of worker productivity and on their risk aversion, the risk aversion of the marginal worker, the curvature of the effort cost function, and the variance of shock on each block. It is straightforward to generalize the analysis above to the more general case where J blocks are available to plant on any given day, and congestion is present (i.e., not all workers can work on their preferred block, given the size of the block and the number of available workers).

4.2.3 Identification

Computation of the costs of risk and the benefits of matching require identification and estimation of a common set of parameters. These include the full distribution of risk aversion parameters among workers within the firm (δ_i) and on the risk levels σ_j. Payroll data on earnings across contracts can be used to estimate the earnings risk and the compensating differentials earned across contracts. However, in the presence of heterogeneous workers, this will identify, at most, the risk preferences of the marginal worker – the worker who is indifferent to working on the different contracts.

To identify the full distribution of risk preferences and the cost of effort, Bellemare and Shearer (2013) combine payroll data with data from two experiments conducted inside the firm. As in the analysis of the profitability of gift-giving presented in the previous subsection, we identify γ using data from the piece-rate experiment analyzed in Paarsch and Shearer (2009). Data from a second experiment are used to identify the distribution of δ_i. This second experiment was inspired by the experimental design exploited by Holt and Laury (2002) to determine the risk preferences of an individual. Each worker was asked to make a sequence of binary decisions between pairs of lotteries that vary with respect to the probabilities of earning specified amounts. Data from the experimental decisions identify a tight bound around a worker's coefficient of risk aversion δ_i. The risk preference distribution can also be used to assess whether sorting on risk preferences is operating in this market. We do so by comparing the estimated distribution of our workers with the corresponding distribution measured using similar instruments with a random

sample of subjects drawn from the Canadian population (see Dave et al., 2008). The table below presents the measured distributions in our experiment and in the broader population for the different interval of δ_i identified by the experimental protocol.

Table
Cumulative Distribution of Risk Aversion

| $U = x^\delta$	Tree-planting firm	Population
$\delta_i > 1.95$	0.085	0.016
$1.49 < \delta_i < 1.95$	0.102	0.021
$1.14 < \delta_i < 1.49$	0.237	0.059
$0.854 < \delta_i < 1.14$	0.458	0.167
$0.589 < \delta_i < 0.854$	0.763	0.295
$0.324 < \delta_i < 0.589$	0.864	0.527
$0.029 < \delta_i < 0.324$	0.915	0.738
$-0.368 < \delta_i < 0.029$	0.966	0.857
$\delta_i < -0.368$	1	1
Sample size	59	881

We find that a significant share of workers in our sample are either risk-neutral or risk-loving. Interestingly, payroll data provide an independent estimate of δ_h, the risk aversion of the marginal worker. The estimated value of δ_h is 2.73, suggesting the marginal worker is risk-loving. Our measured distribution of risk aversion in the firm corroborates the existence of similar risk-loving workers in the firm. Moreover, we find that the proportion of individuals displaying risk-neutral or risk-loving preferences is much lower in the broader population. These results are consistent with the hypothesis that workers match to firms on the basis of their risk preferences: risk-tolerant workers are attracted to high-risk occupations. Our analysis will allow us to predict the benefits of matching with a more representative distribution of risk preferences.

4.2.4 Results

To measure the cost of risk for the workers in this firm, we evaluated (5) at the estimated parameter values for the highest- and the lowest-variance block during the May 2006 planting season. We compute the cost of risk as a proportion of expected earnings. The results are presented in Figure 3.

We find that the costs of risk as a proportion of expected earnings on the high-variance block reveal a similar heterogeneity (left graph): the proportions vary from -15% to 40%, with an average proportion of 1.1%. Unsurprisingly, there is a very small variance in the costs of risks across planters on the low-risk block (right graph), with an average close to zero.

Figure 3

Costs of Risk as Proportion of Earnings
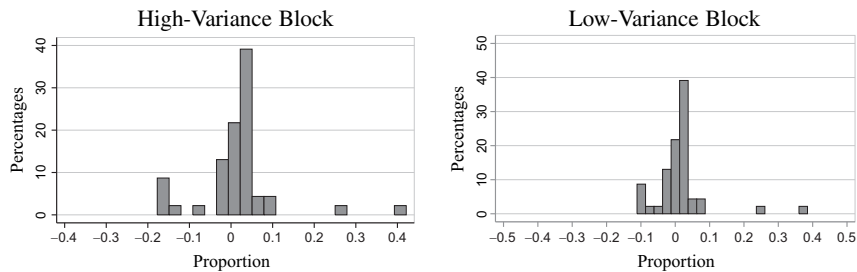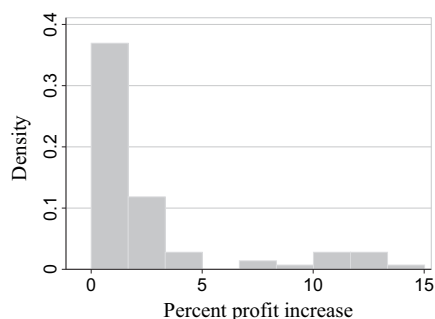

Source: Bellemare and Shearer (2013).

We calculated the potential profits from matching across different planting days throughout the busiest month of the planting season (May) in 2006, taking the available piece rates on each day as given. This involves numerically solving (6), (7), and (8) across the different contracts available on each day.[7]

The results are shown graphically in Figure 4. The average increase in profits is on the order of 2.3 %. This is relatively small in economic terms, yet statistically significant: the bootstrapped 95 % confidence interval is (1.8 %, 3.1 %). This average increase in profits nearly doubles to 4 % if we limit consideration to days on which matching was possible (i.e., days on which more than one block was planted). What is more, for some planting days the profit increase from matching attains 15 %. This suggests that matching can have substantial benefits under certain conditions, although those conditions do not occur very often within this particular firm.

One explanation for the relatively small benefits to matching may be labor-market sorting and the strong presence of risk-tolerant workers in tree planting relative to the share in the broader population. This raises the question of measuring the benefits to matching in an environment where workers' distribution of risk preferences would be more representative. Combining structural estimation with the measured distribution of risk preferences allows us to investigate the importance of sorting in a direct manner. In particular, we simply recompute the benefits of matching, replacing the distribution of risk preferences measured in the firm with the distribution measured for the Canadian population (which is reported in the table above). Overall, we find that the distribution of the returns to matching is very similar to the distribution presented in Figure 4; it is not reproduced here. As a result, the benefits to matching increase only slightly – matching with a more

---

[7] We considered contracts to be blocks of planting paid at the same piece rate within the same geographic location. Our data contained 86 planting-contract days during this month, of which 37 days (43 %) featured only one piece rate and hence no matching possibilities. Of the remaining days, 25 (29 %) featured two contracts, 19 (22 %) featured three contracts, and 5 (6 %) featured four contracts.

Figure 4
Benefits to Matching

Source: Bellemare and Shearer (2013).

representative distribution of risk preferences would increase profits by 4.69 % on days where matching is possible. This represents an increase of less than 1 % relative to the benefits of matching computed using the distribution of risk preferences measured in the firm. It therefore appears that the low benefits of matching measured in this firm are mostly driven by insufficient differences in measured levels of noise σ_j^2 across available blocks on a given day, rather than by the firm attracting relatively more risk-tolerant workers than would be observed in the broader population.

5   Conclusion

Structural estimation and experimentation have become established approaches to analyze a wide range of issues and policies in economics. Despite their natural synergies, few attempts have been made to combine the benefits of both approaches in a single application. Recent applications demonstrated the value of combining both approaches as a way to validate and enrich structural models (e.g., Todd and Wolpin, 2006; Attanasio, Meghir, and Santiago, 2012). The empirical analysis of contracts is an interesting area where complementarities between these approaches are important. A recent illustration is given by Duflo, Hanna, and Ryan (2012), who exploit experimental changes in incentives to estimate the optimal cost-effective incentive scheme limiting teacher absenteeism in India. Contract theory typically involves many unobservables, including effort and preferences. Structural estimation can be used to estimate these effects through specific functional-form restrictions and assumptions on individual behavior. Experimental data provide additional sources of information that enrich the structural approach, reducing the need for modeling assumptions on the underlying unobservables.

Combining both approaches nevertheless requires addressing some important issues. In particular, the experimental design should produce data that can be credibly

used in the context dictated by the model estimated. Experiments conducted with subject pools that differ from the target population under study, or those inducing variations outside the natural operations of a firm, may fail to produce valid data to identify a specific structural model. The approach we propose ideally requires thinking ahead about the model that will be estimated and designing experiments accordingly.

We discussed two applications highlighting these complementarities. The first application demonstrated how structural modeling and experiments can be combined to predict how gift-giving affects worker behavior within the firm and how changes in labor-market conditions affect the profitability of gift exchange. Gift-exchange experiments provide the necessary exogenous variation in compensation needed to identify the reaction of workers to gifts and identify their preferences. The structural analysis complements the experimental variation in that changes in labor-market conditions and outside alternatives are difficult to measure using experimental data. The second application focused on measuring the cost of risk and the benefits to matching. Structural modeling provides a way to characterize these costs and benefits in the firm as a function of effort costs, risk, and risk preferences. They also link these elements to worker reaction to contractual variation, forming a basis of identification. Nonexperimental approaches to identifying these parameters rely on natural variations in contracts, variation that is likely to be endogenous. Experiments offer a natural, direct way to identify these parameters, permitting the calculation of counterfactuals that are based on a limited number of assumptions.

The empirical approach outlined in this paper can be used to analyze other issues in contract theory. For example, recent evidence suggests that workers respond more strongly to wage cuts than to wage increases (see, e.g., Kube, Maréchal, and Puppe, 2013). A common hypothesis used to explain the asymmetric response of workers is preference-based – a worker's marginal utility of punishing the firm for a wage cut is stronger than his/her marginal utility of rewarding the firm for a wage increase of the same magnitude. Field experiments can shed light on this issue, but firms may hesitate to cut the wages of their workers, fearing adverse effects on morale. The structural model of reciprocity discussed in this paper can be used to predict the counterfactual response of workers in the firm to a negative gift. It is straightforward to show that an asymmetric response of workers to a negative gift naturally emerges from the convexity of the marginal cost-of-effort function, despite equating the marginal utility of punishing to the marginal utility of rewarding as measured in this paper (see Bellemare, Marchand, and Shearer, 2015). It follows that observed asymmetric responses to wage cuts and wage increases are not necessarily consistent with the preference-based hypothesis. Additional experimental data with wage cuts could be used, along the lines outlined in this paper, to formally test the preference-based hypothesis.

References

Ackerberg, D. A., and M. Botticini (2002), "Endogeneous Matching and the Empirical Determinants of Contract Form," *Journal of Political Economy*, 110(3), 564–591.

Akerlof, G. A. (1982), "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 97(4), 543–569.

Angrist, J. D., and J.-S. Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton (NJ).

— and — (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *The Journal of Economic Perspectives*, 24(2), 3–30.

Ashenfelter, O. (1987), "The Case for Evaluating Training Programs with Randomized Trials," *Economics of Education Review*, 6(4), 333–338.

Attanasio, O. P., C. Meghir, and A. Santiago (2012), "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA," *The Review of Economic Studies*, 79(1), 37–66.

Bajari, P., and A. Hortaçsu (2005), "Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data," *Journal of Political Economy*, 113(4), 703–741.

Bandiera, O., I. Barankay, and I. Rasul (2007), "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment," *The Quarterly Journal of Economics*, 122(2), 729–773.

Bellemare, C., S. Kröger, and A. van Soest (2008), "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," *Econometrica*, 76(4), 815–839.

—, S. Marchand, and B. Shearer (2015), "Combiner les expériences de terrain et la modélisation structurelle: Le cas de la réciprocité en milieu de travail," *L'Actualité économique*, forthcoming.

— and B. Shearer (2009), "Gift Giving and Worker Productivity: Evidence from a Firm-Level Experiment," *Games and Economic Behavior*, 67(1), 233–244.

— and — (2010), "Sorting, Incentives and Risk Preferences: Evidence from a Field Experiment," *Economics Letters*, 108(3), 345–348.

— and — (2011), "On the Relevance and Composition of Gifts within the Firm: Evidence from Field Experiments," *International Economic Review*, 52(3), 855–882.

— and — (2013), "Multidimensional Heterogeneity and the Economic Importance of Risk and Matching: Evidence from Contractual Data and Field Experiments," *The RAND Journal of Economics*, 44(2), 361–389.

Bolton, P., and M. Dewatripont (2004), *Contract Theory*, The MIT Press, Cambridge (MA).

Brown, M., C. J. Flinn, and A. Schotter (2011), "Real-Time Search in the Laboratory and the Market," *The American Economic Review*, 101(2), 948–974.

Burtless, G. (1995), "The Case for Randomized Field Trials in Economic and Policy Research," *The Journal of Economic Perspectives*, 9(2), 63–84.

Copeland, A., and C. Monnet (2009), "The Welfare Effects of Incentive Schemes," *The Review of Economic Studies*, 76(1), 93–113.

Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2008), "Eliciting Risk Preferences: When Is Simple Better?" available at http://ssrn.com/abstract=1883787.

Duflo, E., R. Hanna, and S. P. Ryan (2012), "Incentives Work: Getting Teachers to Come to School," *The American Economic Review*, 102(4), 1241–1278.

Falk, A. (2007), "Gift Exchange in the Field," *Econometrica*, 75(5), 1501–1511.

Fehr, E., and L. Goette (2007), "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *The American Economic Review,* 97(1), 298–317.

—, G. Kirchsteiger, and A. Riedl (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108(2), 437–459.

Ferrall, C. (2012), "Explaining and Forecasting Results of The Self-Sufficiency Project," *The Review of Economic Studies*, 79(4), 1495–1526.

— and B. Shearer (1999), "Incentives and Transactions Costs within the Firm: Estimating an Agency Model Using Payroll Records," *The Review of Economic Studies*, 66(2), 309–338.

Gneezy, U., and J. A. List (2006), "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74(5), 1365–1384.

Holt, C. A., and S. K. Laury (2002), "Risk Aversion and Incentive Effects," *The American Economic Review*, 92(5), 1644–1655.

Homans, G. C. (1954), "The Cash Posters: A Study of a Group of Working Girls," *American Sociological Review*, 19(6), 724–733.

Keane, M. P. (2010), "Structural vs. Atheoretic Approaches to Econometrics," *Journal of Econometrics*, 156(1), 3–20.

—, P. E. Todd, and K. I. Wolpin (2011), *The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications*, in: O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics: Volume 4A*, North-Holland, Amsterdam, pp. 331–461.

Kube, S., M. A. Maréchal, and C. Puppe (2013), "Do Wage Cuts Damage Work Morale? Evidence from a Natural Field Experiment," *Journal of the European Economic Association*, 11(4), 853–870.

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

Lazear, E. P. (2000), "Performance Pay and Productivity," *The American Economic Review*, 90(5), 1346–1361.

Mauss, M. (1923), "Essai sur le don forme et raison de l'échange dans les sociétés archaïques," *L'Année sociologique: 1896/1897–1924/1925*, 30–186.

Paarsch, H. J., and B. S. Shearer (1999), "The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry," *The Journal of Human Resources (JHR)*, 34(4), 643–667.

— and B. Shearer (2000), "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records," *International Economic Review*, 41(1), 59–92.

— and B. S. Shearer (2009), "The Response to Incentives and Contractual Efficiency: Evidence from a Field Experiment," *European Economic Review*, 53(5), 481–494.

Prendergast, C. (2002), "The Tenuous Trade-Off between Risk and Incentives," *Journal of Political Economy*, 110(5), 1071–1102.

Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics," *The American Economic Review*, 83(5), 1281–1302.

Reiss, P. C., and F. A. Wolak (2007), *Structural Econometric Modeling: Rationales and Examples from Industrial Organization*, in: J. J. Heckman and E. E. Leamer (eds.), *Handbook of Econometrics: Volume 6A*, North-Holland, Amsterdam, pp. 4277–4415.

Salanié, B. (2005), *The Economics of Contracts: A Primer*, second edition, The MIT Press, Cambridge (MA).

Shearer, B. (2004), "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *The Review of Economic Studies*, 71(2), 513–534.

Todd, P. E., and K. I. Wolpin (2006), "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *The American Economic Review*, 96(5), 1384–1417.

Charles Bellemare
Département d'économique
Université Laval
Québec (QC)
G1V 0A6
Canada
cbellemare@ecn.ulaval.ca

Steeve Marchand
Département d'économique
Université Laval
Québec (QC)
G1V 0A6
Canada
steeve.marchand@ecn.ulaval.ca

Bruce Shearer
Département d'économique
Université Laval
Québec (QC)
G1V 0A6
Canada
bruce.shearer@ecn.ulaval.ca

The Complementary Use of Experiments and Field Data to Evaluate Management Practices: The Case of Subjective Performance Evaluations

by

Patrick Kampkötter and Dirk Sliwka[*]

Most firms rely on subjective evaluations by supervisors to assess their employees' performance. This article discusses the implementation of such appraisal processes, exploring the use of multiple research methods such as the analysis of personnel records, survey data, and lab and field experiments to study them in detail. We argue that the complementary use of these methods helps to build a better understanding of how subjective evaluations are conducted and appraisal systems should be designed. (JEL: D22, J33, M12, M52)

1 Introduction

Empirical studies applying different economic methods to evaluate the effect of management practices on firm and employee outcomes have become increasingly important in recent years. In management research, scholars have already for quite some time studied the connection between the use of so-called high-performance work practices (i.e., combinations or bundles of human resource (HR) management practices) and employee or organizational outcomes, typically with quite mixed results (Huselid, 1995; Combs et al., 2006; Subramony, 2009). More recently, also economists have conducted large-scale survey studies to investigate the connection between rather general management practices and firm performance (Bloom and Van Reenen, 2010, 2011; Bloom et al., 2012). Bloom and Van Reenen (2007),

for instance, used telephone interviews to evaluate firms' HR practices along various dimensions, such as monitoring, target setting, and people management, and find that higher "management scores" are correlated with firm performance. An important side effect of these survey studies, sometimes a bit overlooked in the economics literature, is that they give us a broad overview of what firms actually do and how frequently they do it, which in itself is important in that it informs researchers on the relevance of different management practices. But, due to the mainly cross-sectional nature of these data sets, these studies typically cannot establish causal effects. Some of the potential endogeneity issues can be addressed with panel data (Huselid and Becker, 1996; Black and Lynch, 2004), but when there are time-varying unobserved variables that simultaneously affect both the use of a management practice and firm performance, the estimated effects will still be biased.

On the other hand, a recent literature has emerged in economics that focuses on field experiments in firms to evaluate the effect of individual management practices (see, for instance, Bellemare and Shearer, 2009; Bandiera, Barankay, and Rasul, 2011; Englmaier, Roider, and Sunde, 2012; Hossain and List, 2012; Delfgaauw et al., 2013; Manthei and Sliwka, 2014; Friebel et al., 2015). Typically, these field experiments cover a single firm[1] and a specific form of a practice, but in contrast to the broader survey studies they allow a clean and credible identification of causal effects.

Still, for very practical reasons these field experiments are limited in the number of feasible treatment variations, often study very specific industries, and mostly cannot directly observe individual behavior, but only infer information about behavior from performance indicators. Lab experiments, on the other hand, allow for observing behavior directly, and, in principle, researchers can easily implement treatment variations that help to disentangle behavioral channels.

However, the "external validity" of lab evidence is sometimes called into question. If we identify a certain behavioral channel in the lab, it is of course legitimate to ask to what extent we can be sure that this mechanism is of equal importance in a natural setting in a firm and will not be dominated by other mechanisms not captured by the specific experimental design (see, for instance, the discussion in Levitt and List, 2007). Still, as Camerer (2015) argues, the primary goal of most experiments is to understand the general, underlying behavioral principles in a controlled environment and not to establish results that are generalizable from the lab to the field. Nevertheless, he presents ample evidence of lab findings that have proven to be consistent in comparable field settings. But a similar argument can also be made for field experiments: If we have clean causal evidence for a specific result in one firm or a specific industry, to what extent can we be certain that this result will

---

[1] An exception is the field experiment conducted by Bloom et al. (2013) among a large number of Indian textile firms, in which randomly chosen treatment plants received free management consulting services. The results reveal that this informational advantage leads to significant increases in productivity in the treatment firms compared to the control group.

also hold in different firms? In other words, when moving from the lab to a field experiment, we can of course now make a more precise claim about the impact of an intervention on the treated subjects in the specific firm – but we still cannot be sure to what extent this can be generalizable to other firms. Moreover, very often it is simply infeasible to run a field experiment, since for legal or practical reasons certain instruments cannot be randomly assigned. In the context of development economics, Deaton (2010) somewhat provocatively states that the price of success of randomized field experiments in identifying causal effects of a specific program "is a focus that is too narrow and too local to tell us 'what works' " (p. 426).[2]

If we want to collect insights into how management practices affect outcomes in firms, and if we ultimately want to help practitioners to design better management practices, we need to combine the strengths of all these approaches. The key reason is that most management practices affect the performance of organizations through different interlinked behavioral and economic mechanisms. We therefore propose that economic research on management practices should focus on the following three key goals:

(a) to understand the different behavioral mechanisms at work when a practice is applied and to learn how these mechanisms may affect the effects of a management practice,
(b) to estimate the causal effect of its implementation, and
(c) to collect evidence about its relevance and frequency of use in companies.

To achieve these goals, researchers must necessarily apply multiple complementary research methods. Formal economic models help to develop a deeper and more precise understanding of potential behavioral mechanisms. Laboratory experiments are useful to isolate and disentangle these mechanisms in precisely controlled environments.[3] Field experiments and the econometric evaluation of quasi-experiments in firms help us to estimate the causal effect of instrument use on the performance of firms and the well-being of their employees. Moreover, they sometimes even allow us to estimate the magnitude of a performance effect, which can be used as a key ingredient for cost–benefit analyses. Finally, broad representative surveys among firms and employees give us more detailed information about the frequency and correlates of its use in real companies and thus generate further knowledge about the generalizability of insights from lab and field experiments. If there is no widespread adoption of a (well-known) practice for which there is a theoretical

---

[2] Heckman (1992), for instance, has argued that randomization itself could lead to biases in field settings. Imbens (2010) addresses the concerns raised by Deaton (2010) and acknowledges that cases exist where randomization is difficult or not feasible, but he strongly argues that, if they are feasible for the question we are interested in, "randomized experiments are superior to all other designs in terms of credibility" (Imbens, 2010, p. 401).

[3] Ludwig, Kling, and Mullainathan (2011) propose a further distinction stressing the importance of "mechanism experiments," i.e., field experiments that do not directly evaluate the effect of a policy but are designed to study a specific behavioral mechanism underlying a policy.

underpinning and experimental evidence showing that it causally affected performance in a specific environment, we have to ask why this is the case. Hence, in order to get a more comprehensive picture, it is important to exploit the complementary character of these approaches, instead of fighting scientific battles about their relative merits, a point that has already been stressed by Falk and Heckman (2009, p. 537): "[...] empirical methods and data sources are complements, not substitutes. Field data, survey data, and experiments, both lab and field, as well as standard econometric methods, can all improve the state of knowledge in the social sciences. There is no hierarchy among these methods, and the issue of generalizability of results is universal to all of them."

In this paper, we want to illustrate and discuss potential applications of this mix of complementary methods, using the example of one important HR management practice, namely subjective performance evaluations of employees (for other applications, see Englmaier and Schüßler, 2016). Subjective performance evaluations are widely used in many firms, but we still do not precisely know whether and how they affect various outcomes such as job performance or the satisfaction of employees. In detail, we present the results of several of our own studies using various methods to analyze the influence of differentiation in performance evaluations on the provision of individual efforts, as well as employee perceptions like job satisfaction or fairness perceptions in response to performance appraisals. Examples include the use of linked employer–employee data, an industry-wide field study, and an *insider econometrics* approach combining data from personnel records and employee surveys as well as a field and laboratory experiment.

2 Purposes of Performance Evaluations and Design Challenges

2.1 Purposes

Performance evaluations, or performance appraisals,[4] are mainly used to evaluate and monitor the contributions of individual employees to overall firm performance. They often combine the use of objective performance indicators and subjective evaluations. Firms typically use performance evaluations for multiple reasons (Landy and Farr, 1983; Murphy and Cleveland, 1991, 1995). First, in most incentive schemes individual bonuses are based on subjective and objective performance indicators generated through appraisal processes. Second, performance evaluations are often the starting point for employee *development* decisions, such as the assignment of training. Third, appraisal outcomes are used in the personnel planning process for decisions on promotions, reallocations, or dismissals. Frederiksen, Lange, and Kriechel (2012), for instance, analyze data sets on subjective performance ratings from six large international companies that have been used in several prominent studies on internal labor markets; they find that performance

---

[4] In the following, "performance appraisals," "performance assessments," and "performance evaluations" are used interchangeably.

evaluations predict career outcomes such as promotions (see also Halse, Smeets, and Warzynski, 2011). Fourth, performance appraisals give employees direct feedback about their performance and potential strengths and weaknesses. Feedback can show employees whether to reallocate efforts or to invest in new skills and, moreover, can have a direct impact on employee satisfaction and thus the decision to stay with an employer (Fletcher and Williams, 1996; Whitman, Van Rooy, and Viswesvaran, 2010; Kampkötter, 2016).

2.2 Design Challenges and Appraisal Formats

While performance appraisals sometimes also include objective performance information (such as financial key figures), most often subjective assessments by a supervisor play a dominant role. A key reason for this is that in many cases objective indicators of individual performance are not available (an exception is, for instance, the sales function, where objective performance measures are nearly always available and frequently used). This is typically the case in many cross-functional positions, such as human resources, controlling, finance, and marketing. Objectively measurable performance indicators can often only be derived jointly at the team level, and individual contributions to this team output are difficult to evaluate. Furthermore, individual performance is frequently rather complex and cannot be tracked with a small set of performance indicators. When individual performance strongly depends on external factors that are outside the control of employees (the market situation, the state of the economy, etc.), objective performance measures can only be crude indicators for employees' efforts and talents. As a result, the majority of performance evaluations in practice are based on subjective assessments by supervisors.

There is substantial evidence, mostly from research in personnel psychology, showing that these subjective evaluations are typically biased. Firms commonly use systems in which employees are assessed on a given scale (for instance, with evaluation grades ranging from 1 to 5), and often only a subset of the scale is actually used by supervisors. Psychologists have coined the terms *centrality* and *leniency bias* to describe patterns that are frequently observed (Landy and Farr, 1980; Murphy, 1992; Prendergast and Topel, 1993; Kane et al., 1995; Murphy and Cleveland, 1995; Prendergast, 1999; Gibbs et al., 2004; Moers, 2005; Frederiksen, Lange, and Kriechel, 2012). In the case of a centrality bias, the variation in performance appraisals is smaller than intended by the designer of the system, i.e., supervisors do not use the full range of the rating scale and particularly avoid marginal grades. The so-called leniency bias describes a phenomenon where the mean of the appraisal ratings is higher than the mean of ratings intended by the firm, i.e., supervisors systematically evaluate their subordinates more highly than they are supposed to. Additionally, supervisors differ in the extent to which they are prone to these biases. Heterogeneity in the supervisors' types therefore leads to heterogeneous evaluation behavior even within the same firm (see, for instance,

Bernardin, Cooke, and Villanova, 2000, for a study on the role of personality factors in appraisal behavior).

In practice, firms have adopted several instruments to reduce potential biases that presumably lead to more differentiation among employees. Two instruments prominently used in practice are *recommended distributions* and *evaluation panels*. When a firm adopts a recommended distribution, it tells managers the relative proportion of different grades that should be assigned. In the appraisal system of the multinational firm studied in Ockenfels, Sliwka, and Werner (2015), employees were rated on a 5-point scale, and the firm recommended the following distribution:

Grade 1 ("excellent"): $\leq 5\%$ of employees;
Grade 2 ("above average"): $\leq 25\%$ of employees;
Grade 3 ("fully meets expectations"): $\approx 60\%$ of employees;
Grade 4 ("below average") and grade 5 ("inadequate"): together $\leq 10\%$ of employees.

However, as recommended distributions are nonbinding and just give a guideline on how to assess performance, they may be accompanied by more lenient and less differentiated actual ratings. In the studied firm, for instance, ratings were tied to (budgeted) bonus payments; more than 65 % of employees received a rating of 3, and more than 30 % a rating of 2. On the other hand, less than 5 % received a 4, and nearly nobody a 5, which is also a very common occurrence in other firms. Frederiksen, Lange, and Kriechel (2012), for instance, investigate typical patterns in subjective evaluations from several data sets used in the prior literature and find similar or even more extreme patterns.[5]

Some firms therefore adopt stricter so-called *forced distributions*, where these proportions are not guidelines, but rather the appraisal process is designed so that the evaluators must adhere to a given exact distribution. The most prominent example of such "grading on a curve" is General Electric, where Jack Welch's "vitality curve" forces managers to identify the top 20 % and bottom 10 % of employees each year (see Welch and Byrne, 2001, chapter 11).

A key challenge in grading employees is that individual managers often supervise and therefore evaluate only a small set of employees. Even if a manager wants to be accurate in rating her employees and is able to rank them, the fact that somebody belongs to the top 20 % in a certain unit does not guarantee that this person belongs to the top 20 % of the whole firm. If the mix of talents and performance is unequally distributed across teams, this either directly leads to different evaluation standards (if managers have to stick to the distribution within their team) or

---

[5] Among white-collar employees of the former Dutch airplane manufacturer Fokker (Dohmen, Kriechel, and Pfann, 2004), 81 % received the middle grade and 14 % the top two grades. In the Baker–Gibbs–Holmstrom data set (Baker, Gibbs, and Holmstrom, 1994a,b) of a U.S.-based service-sector firm, 82 % of the employees were rated with one of the two top grades, and in the Flabbi–Ichino data set (Flabbi and Ichino, 2001) of a large Italian bank, this fraction is 83 %. Note that in all of these examples, 5-point rating scales are applied. See Frederiksen, Lange, and Kriechel (2012) for details.

makes the process very complicated in that evaluations have to be coordinated (if managers try to adhere to the distribution not within each team, but across a larger number of teams). As a response, many bigger firms have in recent years adopted so-called *evaluation panels* (sometimes also called calibration meetings, management panels, or evaluation roundtables), in which a group of managers meet to discuss the performance evaluation of all of their employees. It is quite common, for instance, that a group of 60 or 80 employees are discussed in such a panel and that top management and HR representatives are involved in this process.[6] These panels serve to "calibrate" evaluations made by individual managers in order to generate common standards. Moreover, if a recommended or forced distribution is adapted, these panels make it easier to stick to this distribution, reducing the likelihood that employees receive unjustified ratings because of the composition of their direct work group.[7]

In section 3, we present recent descriptive evidence of the use of these appraisal procedures from a novel representative data set on HR practices.

2.3   The Controversial Role of Differentiation

From an economics perspective, subjective evaluations entail a potential conflict of interest between the evaluating supervisor's personal interest and the interests of the firm in its role as the employer of both the supervisor and the evaluated employee. A large body of evidence in behavioral economics has shown that people have social preferences (see Fehr and Schmidt, 2006, for a survey), i.e., their own well-being also depends upon the well-being of other people in their proximity. In this respect, there is evidence that both direct *altruism* (i.e., a person can ceteris paribus be better off when another person has a higher payoff) and *equity concerns* (i.e., a person is ceteris paribus better off when an outcome leads to a more equal payoff distribution) matter.[8] Moreover, in both respects *reference points* seem to play an important role in light of the substantial evidence that people often evaluate outcomes relative to a reference standard (Kahneman and Tversky, 1979). And these reference standards are affected, for instance, by people's own prior expectations (Bell, 1985; Loomes and Sugden, 1986; Kőszegi and Rabin, 2006) or by *social reference points,* i.e., the outcomes of others (Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

---

[6] See, for instance, Michaels, Handfield-Jones, and Axelrod (2001) or Welch and Byrne (2001) for a description of typical processes.

[7] To give an example: If a manager has to evaluate five direct subordinates and can assign the highest evaluation to the top 20 %, the likelihood that either no one or more than one of her direct subordinates belong to the top 20 % in the firm is rather large. The bigger the group, the smaller is the likelihood that such unjustified ratings have to be assigned.

[8] But behavioral economics research has also established that individuals may have not only prosocial but also antisocial concerns such as spite or envy, or even direct antisocial preferences (see, for instance, Zizzo and Oswald, 2001, or Abbink and Sadrieh, 2009).

If now a supervisor has to evaluate a direct subordinate, she has to trade off the effects of her rating on the employee's well-being and the firm's performance. Hence, supervisors' preferences will often not be perfectly aligned with the interests of firms. For example, while an employer may prefer to have accurate evaluations that reflect differences in performance, supervisors may be tempted to assign generous or ungenerous ratings to discriminate between their subordinates. Very often, subjective evaluations determine bonus payments to employees. Assigning better ratings thus leads to higher wages, often without substantial direct costs for supervisors.[9] A supervisor who cares about the well-being of her subordinate will therefore have a tendency to assign more lenient ratings (see Prendergast and Topel, 1996, or Giebe and Gürtler, 2012, for formal models on this issue). Breuer, Nieken, and Sliwka (2013), for instance, analyze longitudinal data from a call-center organization where objective performance measures are available. They use variation in team composition over time to show that employees receive better ratings for the same objective performance when they have worked with the same supervisor over a longer time or when the supervisor manages a smaller team, demonstrating the effect of social ties between supervisors and subordinates on appraisals. Supervisors may also avoid poor ratings out of a reluctance to provide negative feedback, even if the actual performance was poor, because negative feedback typically has to be justified in more detail and may induce "psychological costs." Lenient ratings will also more likely prevent conflicts with subordinates (Varma, Denisi, and Peters, 1996). On the other hand, the social preferences of a supervisor may also affect the scope for relational contracts if supervisors and subordinates interact repeatedly (see Dur and Tichem, 2015, for a formal analysis of the role of a supervisor's altruism but also potential spitefulness in relational contracting under subjective evaluations).

Frequently, bonus payments have to be paid from a given budget; this is particularly true in the banking and financial services sector (Kampkötter and Sliwka, 2014). In this case, rater "leniency" is restricted, as there is an upper limit on the assignable ratings. But when supervisors either have a preference for equity among employees or take employees' equity concerns into account, a "bias towards centrality" or reduced *differentiation* among employees directly follows.[10]

On the other hand, even if ratings assigned by supervisors are not fully accurate, it is not clear at the outset whether an apparent lack of differentiation could in fact be due to reasonable behavior by supervisors that even may be to some extent in line with the firm's interest. Appraisal patterns interpreted as a bias from one perspective may in principle be beneficial from another perspective. Consider the following example. Suppose that we are looking at a system in which supervisors are asked to give the worst 20 % of employees the worst performance grade. A su-

---

[9] Sometimes there are of course indirect costs (as will be made clear below): for instance, employees' performance may be lower, which in turn may hurt the supervisor.

[10] See Grund and Przemeck (2012), Kampkötter and Sliwka (2014), or Ockenfels, Sliwka, and Werner (2015, online appendix) for formal models analyzing the role of supervisors' or subordinates' preferences for equity in performance appraisals.

pervisor who does not assign this worst grade to an employee he considers to be actually in the bottom 20 % of course creates a bias if the purpose of the rating is to identify the employee's relative standing in the talent distribution. Moreover, from a neoclassical incentive perspective, such a bias may also be detrimental, as low performance is then not penalized and high performance not adequately rewarded, which may reduce the incentive to exert higher efforts in the future. However, this supervisor may argue that his rating behavior is justified by another purpose. A large literature in experimental economics, starting with Fehr, Kirchsteiger, and Riedl (1993) and Fehr, Gächter, and Kirchsteiger (1997), has established that individuals have a preference for reciprocity. Thus supervisors' leniency may actually to some extent act as a trigger for a higher employee motivation through positive reciprocity, or the avoidance of low ratings may arise from the fear of demotivating agents and causing negative reciprocity. Sebald and Walzl (2014), for instance, show in a laboratory experiment that employees punish their supervisors when assessments deviate negatively from the employee's self-evaluation of her own performance.[11] Hence, there may be trade-offs between the accuracy of the ratings and its other purposes. Even if we only take an incentive perspective, there may be a trade-off between triggering more social motives via positive reciprocity or extrinsic motives by punishing low performance. On top of that, excessive leniency may lead to unfair treatment of high performers and reduce their motivation: If high performers have a concern for equity not only in bonus payments but also in the exerted effort costs, they may be tempted to reduce their efforts if low performers receive generous ratings at lower effort levels. This highlights the importance of studying these trade-offs in detail. In the subsequent sections, we will first show descriptive evidence on the use of performance appraisals and practices to foster differentiation, before presenting a number of different studies that apply different methods to study in greater detail the role and influence of differentiation in performance evaluations.

3   The Use of Performance Evaluations in Firms: Descriptive Evidence

Performance appraisals are a core element of personnel policies in most firms. However, firms differ in the extent to which they apply appraisals. We start by analyzing early evidence from a new, representative linked employer–employee data set of German firms, the *Linked Personnel Panel (LPP)*, which contains information about key elements of the appraisal process.[12] The authors designed this sur-

---

[11] Interestingly, the negative reciprocal action of employees to this perceived unkind act by their supervisors also holds if the appraisals have no monetary consequences. See also Takahashi et al. (2014), who analyze the personnel records of sales representatives in a major Japanese car sales company and show that measures of rating biases are positively related to employee quits.

[12] See Bellmann et al. (2015) or Kampkötter et al. (2016) for an overview on the structure of the data set.

vey jointly with the Centre for European Economic Research (ZEW) Mannheim as part of a project on behalf of the German Federal Ministry of Labor and Social Affairs and the research institute of the Federal Employment Agency (IAB). So far, the first wave of both the employer and employee survey has been administered in 2012 and 2013, and the second wave will be completed in late 2015. The representative firm-level survey includes 1,219 private-sector establishments employing at least 50 employees. The employee survey includes 7,508 randomly drawn respondents from 869 of these establishments. The key aim of this new data set is to develop a longitudinal infrastructure to assess the effect of HR practices on employees' *quality of work* (e.g., satisfaction, engagement, and turnover) and the economic success of firms. The LPP links employer-level information about HR policies with employee-level information about attitudes and behavior, and enables researchers to analyze how individuals perceive and respond to HR policies. The data set provides information on various HR instruments on the firm level, including dimensions such as recruiting, performance management, employee and career development, training, corporate culture, and promotion of female employees. The employee questionnaire mirrors some of these practices, such as training, promotion, and career development, and additionally elicits information on employee perceptions such as job satisfaction, commitment, fairness perceptions, risk attitude, and personality traits. Over time, the survey will evolve into a panel data set that will allow one to study within-firm variation of HR practices and link this to potential changes in employee perceptions and firm performance. Currently, the available data from the first cross section yields descriptive evidence on specific practices applied in performance evaluations.

Figure 1 shows the frequency of use of performance appraisals[13] by establishment size. We find that the majority of establishments use structured performance appraisals, with the frequency of use increasing from 62 % in 2012 to 67 % in 2014 across all establishment sizes. As the figure shows, larger establishments use systematic appraisals more often than smaller ones. However, we observe the most substantial increase in the smallest establishments.

In a second step, we also asked about the appraisals' target group, for example, whether the practice is applied only to employees in a supervisory role (i.e., managers) or to all employees. As Figure 2 shows, a majority of the establishments that use systematic appraisals indeed use these appraisals for all employees, and this fraction has (slightly) increased over the short time frame we consider.

In the survey we also asked firms about the use of *recommended distributions* and *evaluation panels* (joint evaluation by more than one supervisor).[14] Figure 3

---

[13] The exact item is as follows: "Is the performance of employees in your establishment evaluated by supervisors at least once in a year?"

[14] The survey items are "Does a recommended distribution for performance assessments exist in your establishment? Recommended distributions convey information about the proportion of employees who should receive the best rating, the second-best rating, etc." and "Are employees typically assessed by *one* supervisor or jointly by *a group of supervisors* (management panels), i.e. not only by one supervisor?"

Figure 1

The Use of Performance Appraisals by Establishment Size
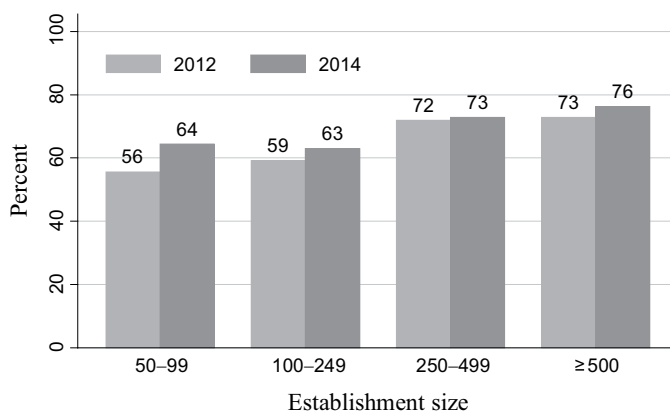
Notes: N = 1217 and 771.
Source: LPP establishment survey waves 2012 and 2014.

Figure 2

Employee Target Groups of Performance Appraisals
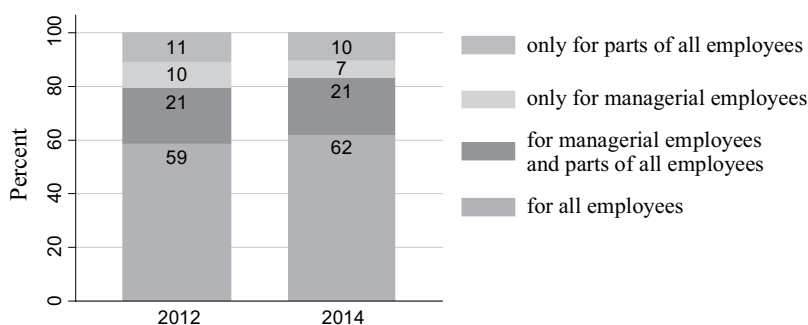
Notes: N = 745 and 517; restriction: only establishments that conduct performance appraisals at least once a year.
Source: LPP establishment survey waves 2012 and 2014.

shows the distribution of the intensity of use over time for both practices. In 2012 only about 15 %, and in 2014 only 13 %,[15] of the establishments using performance appraisals employed recommended distributions. However, the frequency of their use is higher in larger establishments: in establishments with 50–99 employees, only 10 % employed recommended distributions, but this increases to 23 % in establishments with more than 500 employees.

_____

[15] We note that changes over time are partially due to sampling of firms, as the LPP is an unbalanced panel.

Figure 3
Recommended Distributions and Evaluation Panels

Notes: $N = 748$ and 517 (recommended distributions) and
731 and 513 (evaluation panels) per wave; restriction: only establishments that conduct performance appraisals at least once a year.
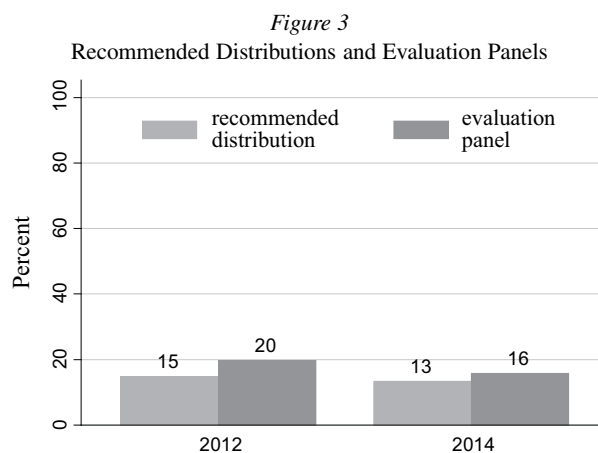*Source:* LPP establishment survey waves 2012 and 2014.

In 2012 about 20 %, and in 2014 about 16 %, of the establishments using performance appraisals used evaluation panels or based their appraisals on joint evaluations by more than one supervisor. It is interesting to observe that this is more frequent in very small establishments (24 % in 2012), less so in medium-size establishments (13 %), and again more so in establishments with more than 500 employees (22 %). This could possibly stem from the fact that in small firms the management team talks about the assessment of all their employees more often, as the likelihood is sufficiently large that all members of the management team know all employees. This becomes more difficult in larger establishments. The larger the establishment becomes, the larger is then the likelihood that formalized procedures are used that establish a structure for group evaluations in panel discussions.[16]

The overall frequency of appraisal use shown in the above is in line with survey evidence from individual employees. Kampkötter (2016) uses the German Socio Economic Panel (SOEP), a large and representative panel data set of a subset of the German population,[17] to investigate the intensity of the use of performance appraisals among employees and their effect on job satisfaction. In the years 2004, 2008, and 2011, individual employees were asked whether their performance is evaluated regularly by their supervisor and whether these appraisals have monetary

---

[16] In fact, as our own discussions with HR managers of several big DAX30 companies in Germany have shown, structured evaluation panels are a typical element of their appraisal procedures, and often there are tight guidelines regulating their implementation.

[17] See also Grund and Sliwka (2009) for an earlier study on a related issue using cross-sectional data from the SOEP.

(impact on gross salary, annual bonus, wage increase, promotion) or nonmonetary consequences. Descriptive results reveal a positive trend: only 32 % of all employees in the sample were subject to systematic performance appraisals in 2004, and this number increased to 39 % in 2011.[18] An even stronger increase of more than 50 % can be found for appraisals linked to individual bonus payments (from 11 % in 2004 to 17 % in 2011).

4   The Use of Field Data and Experiments to Evaluate the Effects of Differentiation among Employees

As the previous section has shown, subjective performance evaluations are a core element of firms' HR practices. But there is little empirical evidence on the incentive effects of performance appraisals, as noted by Rynes, Gerhart, and Parks (2005), who state in their survey that "although there is a voluminous psychological literature on performance evaluation, surprisingly little of this research examines the consequences of linking pay to evaluated performance in work settings" (p. 571). We have also seen that there are important trade-offs in the design of appraisals. In particular, whether driving supervisors to assign more differentiated ratings is indeed beneficial for performance represents a very important question for the design of appraisal systems in practice. In the following, we discuss different aspects of differentiation in subjective evaluations, using as examples our own studies applying different methods, comprising an industry-wide field study, an insider econometrics approach combining personnel and survey data, a laboratory experiment, the use of linked employer–employee data, and a field experiment.

4.1   Incentive Effects of Performance Appraisals: An Industry-Wide Field Study

From a typical agency perspective, where a moral-hazard problem has to be solved, performance appraisals should be structured in such a way that high performance is rewarded and low performance penalized adequately. Biases in performance appraisals may therefore weaken the incentive effect, because the relationship between actual efforts and assigned ratings – i.e., the marginal return an employee gets from one unit more effort – decreases with increasing magnitude of the bias. However, as argued in the above, there may be countervailing behavioral effects, as low ratings may also trigger negative reciprocity. In the already mentioned lab experiment by Sebald and Walzl (2014), agents had the opportunity to reduce the principal's payoff, at a cost, in response to the feedback provided by the super-

---

[18] Note that this is not the percentage of establishments using structured performance appraisals as in the firm-level data presented in the above, but rather the percentage of employees among a representative selection of employees in Germany. The fact that the fraction of employees with performance appraisals is smaller than the fraction of establishments using it is due to several factors. First, as laid out in the above, not all firms use it for all employees. Moreover, the frequency of use is lower in small establishments, and the smallest establishments (with less than 50 employees) are not part of the LPP survey.

visor, and they actively made use of this option. In firms, negative reciprocity, i.e., punishing a supervisor, might manifest itself in different ways, for instance, by not providing sufficient effort, being absent from work, disturbing the working climate in the unit, or badmouthing or even sabotaging the supervisor.

Only a small number of studies have so far investigated the relationship between differentiation and indicators of performance in field settings, with nearly all of them looking at single firms. Bol (2011) analyzes a longitudinal sample of 200 employees working in five branch offices of a Dutch financial services firm. She finds a positive relationship between a higher differentiation in performance ratings (i.e., a reduced centrality bias) and subsequent objective performance indicators in around 35 teams each year. Engellandt and Riphahn (2011) also study personnel records from one firm, showing that a higher dispersion in performance ratings is positively associated with higher performance in the future, proxied by paid and unpaid overtime.[19]

In Kampkötter and Sliwka (2014), we analyze the impact of differentiation on subsequent performance in an industry-wide field study. We make use of a panel data set on compensation in about 40 German banks provided by an international management consultancy. The data set provides information on fixed salaries, short-term performance-related bonus payments, age, firm tenure, hierarchical level (six levels), functional areas (eight areas), career ladder (management and expert positions), and specific functions (about 60 functions) for the years 2005–2007. As a complementary survey study illustrates, the bonus schemes used in these banks are frequently so-called *bonus pool* arrangements, whereby the bank allocates a sum of money to individual units, which is then distributed to the employees, mainly according to subjective performance evaluations. The size of the bonus pool is typically a function of the financial success of the unit. The key idea of the empirical approach is to estimate the causal effect of differentiation in bonus payments within a unit on the size of the bonus pool in the subsequent year, which should reflect the financial success of this unit. In other words, the question is: does within-unit differentiation in bonus payments affect the success of a unit? To estimate the degree of differentiation, work units are identified by a unique combination of year, company, function, ladder, and hierarchical level. In a next step, the coefficient of variation in bonus payments is calculated for each work unit and year.

In the main specifications, fixed-effects models are estimated, where individual bonus payments in a period t are regressed on the work-unit-level measure of differentiation (coefficient of variation) in the previous period $t-1$. The results show a positive and statistically significant average effect of a within-work-unit change in differentiation on subsequent individual performance of employees. To evaluate the economic significance of this incentive effect, the degree of differentiation is

---

[19] Regular overtime work was not remunerated financially, but used to substitute for working hours (paid overtime). As employees were not allowed to carry more than 120 overtime hours from one month to the next, those employees having accumulated more than 120 overtime hours provided free labor to the company (unpaid overtime).

divided into quintiles. Moving from a work unit that belongs to the 20 % weakest "differentiators" to a work unit that belongs to the top 20 % with respect to the degree of differentiation comes along with an increase in subsequent bonus payments by more than 30 %. The results are qualitatively robust in instrumental-variable regressions to allow for the potential endogeneity of changes in the degree of differentiation. In particular, we try to identify factors affecting the dispersion in a unit that are exogenous to this unit's performance. We construct an instrument that measures the average degree of differentiation of other work units within a functional area of the same company and hierarchical level (excluding the work unit we are looking at) for each year. Changes, for instance, in a firm's general evaluation policies and guidelines should affect all departments in a company and therefore be reflected in this instrument. The identifying assumption is that the level of differentiation in other units does not have a direct effect on the bonus payments in a particular area, beyond the influence through the dispersion in the area itself.

Further analyses on subsamples show that the effect is the strongest at the intermediate and highest hierarchical levels. However, the picture changes at the lowest levels, where more differentiation is even associated with lower subsequent average bonus payments. Moreover, there are differences among functions, and there is evidence that differentiation works better in functions where performance evaluation is less subjective (such as retail banking). It is argued in a formal model that a lack of willingness to differentiate is more detrimental with more objective evaluations, where the potential loss in achievable extrinsic incentives is the largest. To better understand potential drivers of detrimental effects of differentiation, it is thus useful to dig deeper into the behavioral processes underlying the link between appraisals and employee behavior.

4.2 *Forced Distribution and Performance: A Lab Experiment*

Berger, Harbring, and Sliwka (2013) analyze the effect of a forced distribution system on the differentiation of performance ratings in a controlled lab experiment. Of course, as stated above, we are sympathetic to the view that researchers should be cautious when deriving direct practical implications, as a lab experiment can never fully capture the richness of a real-world employment relationship. However, lab experiments are uniquely suited to disentangle behavioral mechanisms by intentional design of different treatment variations and to measure specific individual reactions.

In the experiment, participants were assigned to fixed groups of three *workers* and one *supervisor*. Workers had to work on a real-effort task for eight rounds in the main part of the experiment.[20] The timing in each round was as follows: First, workers had to perform a tedious real-effort task. Afterwards, workers and the supervisor learned the performance of all group members. The supervisors of

---

[20] Prior to the treatment intervention, all subjects had to work on the same task for a piece rate in order to obtain a measure of their ability. Workers were matched based on this ability measure in order to have homogeneous groups.

each group then rated each worker on a 1–5 scale (with 1 being the best grade and 5 the worst), and workers were privately informed about their own rating at the end of each round. Performance ratings determined expected bonus payments paid to workers.[21]

In the main treatments of the experiment, bonuses were not paid by supervisors (as is common in most firms, where supervisors are typically not the owners of the firm). The supervisors' own payoff was a linear function of the workers' performance on the task, so that supervisors had some interest in evaluating workers in a way that increases performance. But of course, they may have also directly cared about the well-being of the participants in the role of workers.

The main experiment consisted of two treatments. In the baseline treatment, supervisors faced no restriction on how to assign ratings to their workers. This is compared with a forced distribution treatment, where it became mandatory for the supervisor to rate one worker with a grade of 1 or 2, one worker with a 3, and one worker with a grade of 4 or 5. Rating distributions in the baseline treatment show strong evidence for rater leniency, as more than 80 % of all workers received a 1 or a 2, whereas less than 10 % of all workers were given a 4 or 5.[22]

Interestingly, as a postexperimental elicitation of preferences shows, more altruistic subjects assign more lenient ratings and more equity-oriented subjects assign less differentiated ratings, which is well in line with the idea that social preferences affect rating behavior. Moreover, it shows that heterogeneity in supervisors' types is an important element that has to be considered when appraisal systems are designed.

The main result of the experiment is that the forced distribution raised group output significantly, by about 6 % in the main experiment and by about 12 % in a set of treatments where supervisors had to share the costs of the bonus payments. Further analyses of direct effort reactions to grading reveal that leniency indeed reduced performance. Hence, potential effects of positive reciprocity were apparently dominated by extrinsic incentive effects: When workers realized that they could earn high bonuses even without working harder, they apparently reduced their efforts. This was not possible in the forced distribution treatment, in which there was ongoing competition for the high grades.

But the forced distribution also had detrimental effects, as shown in a further treatment variation. In the third treatment, subjects had access to an option whereby they could anonymously block their coworkers' computer screen for 20 seconds. This sabotage effort was costly in that subjects' own screens were also blocked for 3 seconds. In this treatment group, output is significantly lower when a forced distribution is employed. Hence, when cooperation among employees is impor-

---

[21] In the experiment bonuses were awarded for each period, with a bonus of €10 for the highest rating and €0 for the lowest. One of the rounds was randomly drawn at the end of the experiment and paid out.

[22] Implementing the forced distribution of course reduced leniency, but supervisors still had the discretion whether to give one worker a 1 or a 2 and one worker a 4 or a 5. The vast majority of workers were rated 1 in the first case and 4 in the latter.
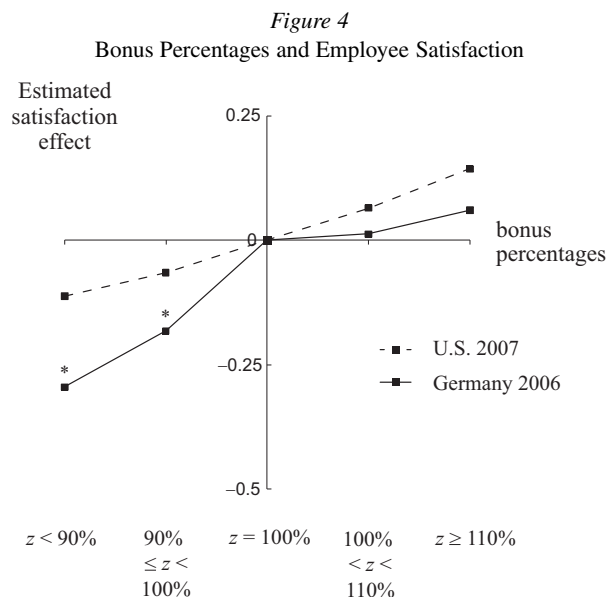
tant (or sabotage is easy), a culture of higher (forced) differentiation may lead colleagues to become competitors, creating negative side effects. This yields one potential explanation for the observation in Kampkötter and Sliwka (2014) that differentiation can be harmful at lower hierarchical levels. At these levels, employees in similar jobs are often direct colleagues who share offices. Hence, in these jobs competition may harm more than it helps performance.

4.3 Evaluations and Reference-Point Violations: Combining Personnel Records and Survey Data

As laid out above, social preferences are important drivers of human behavior and thus should affect subjective evaluations. Moreover, the perception of and reaction to subjective evaluations should be affected by reference standards, such as employees' prior expectations and social comparisons with the outcomes of others. Ockenfels, Sliwka, and Werner (2015) study this in detail, investigating the bonus and appraisal scheme for managers of a multinational company.

The study combines detailed data on performance evaluations from personnel records with survey data on employee perceptions. The panel data set on performance evaluations comprises information on compensation and bonus payments of all (several thousand) managers of the company for Germany (2004–2006) and the United States (2004–2007). This data is supplemented by an anonymous employee survey among managers, eliciting, for instance, their job satisfaction, which is then matched to the appraisal data on the individual level. The study thus follows an insider econometrics approach (see, e.g., Bartel, Ichniowski, and Shaw, 2004) combining the econometric analysis of firm data with detailed institutional knowledge about the firm and survey evidence. The employee survey was administered to German managers in autumn 2007 and to managers in the United States in summer 2008.

The bonus and appraisal system in this company was organized in the following way: Each year, managers were rated by their respective supervisors on a 5-point rating scale. Several weeks later, supervisors had to assign bonus payments to the managers. For each manager an individual *bonus budget* was determined, which depended on the manager's salary grade as well as the performance of the company and the respective division. Hence, when having the same salary grade, all managers in a unit also had the same bonus budget. Supervisors then determined individual bonuses, subject to the constraints that (i) the sum of bonuses did not exceed the sum of bonus budgets and (ii) bonus payments reflected the previously assigned performance grades. For instance, a manager with a rating of "Fully meets expectations" had to receive a bonus between 80 % and 110 % of the budget assigned to him. Better-rated managers had to be assigned at least 110 %, and worse-rated managers less than 80 %, of the budget. While the rules were otherwise identical in Germany and the U.S., there was one key difference: In Germany, managers learned not only the amount of their bonus, but also the payout percentage, i.e., what percentage of the budget allocated to them they actually received.

Figure 4
Bonus Percentages and Employee Satisfaction

Source: Ockenfels, Sliwka, and Werner (2015).

In the U.S., managers were only informed about the absolute amount of the bonus and were not told the payout percentage. A key difference between the system in Germany and the U.S. is thus that in Germany managers could directly compare their bonus with (i) the average bonus in their team and (ii) their own prior expectations based on the assigned performance grade. It is now argued that a payout percentage of 100 % is an important reference standard. Falling below 100 % implies, for instance, that a manager received less than the average of her team. When inequity aversion plays a role, this should be accompanied by a utility loss beyond the monetary consequences.

The relationship between job satisfaction and absolute bonus payments and payout percentages is then analyzed. Figure 4 illustrates the key result. It shows the coefficients of a simple OLS regression with a unit-normal transformation of the satisfaction score for Germany and the U.S., including dummies for intervals of the bonus percentages. The reference group consists of managers who receive exactly 100 %. Hence, the graph normalizes satisfaction at the level of managers with 100 % bonus in each country and displays the satisfaction effect of managers with other bonus percentages relative to this benchmark. In the German sample, both interval dummies below 100 % are significantly smaller than zero. Both interval dummies above 100 % are positive but statistically insignificant. In the U.S., none of the dummies is statistically different from zero.

Hence, bonus payments below 100 % reduce employee satisfaction in Germany, where the system creates a salient comparison standard, but not in the U.S. Em-

ployees seem to use the target bonus of 100 % as a reference point, and negative deviations from this reference point have a stronger impact on their well-being than positive deviations. Reference-point violations here most likely have this strong effect because a bonus below 100 % leads to a violation not only of expectations but also of a social reference standard, as they reveal that a manager gets less than her colleagues.

Further analyses detect indications of a negative performance effect: supervisors who create more reference-point violations among their subordinates themselves attain a lower performance rating in the subsequent year.[23] A complementary lab experiment replicating qualitative features of the studied environment shows that salient reference-point violations trigger negative reciprocal reactions towards supervisors.

The tension between potentially positive incentive effects of differentiation described in the previous sections and the potentially negative effect in that differentiation may frequently come along with reference-point violations is an interesting point to discuss. One insight is that fine-grained differentiation can be detrimental when performance evaluation is subjective. In the discussed study, the negative effects were basically driven by managers who were all rated as "fully meets expectations," but some received a bonus of 100 % of their budget, while others, for instance, received only 96 %. In monetary terms these are small differences, but they have a substantial effect on well-being. Here it is quite likely that a rather affective negative reaction to reference-point violations may outweigh potential positive effort effects through higher-powered incentives because the latter effect is weak. Indeed, a further analysis of the data finds no indication that differentiation across grades (i.e., the 1–5 performance ratings) is detrimental. It apparently is the differentiation within a grade, i.e., of managers with very similar performance levels, that is problematic for satisfaction and, in turn, performance.

4.4 Objective Performance Measurement: A Field Experiment

As already argued in section 4.1, objective performance information may help to facilitate differentiation and foster incentives. Manthei and Sliwka (2014) investigate a field experiment on the benefits of objective performance measurement in performance evaluations. A retail bank in Germany conducted the field experiment in 2003 in order to evaluate the causal effect of the use of objective performance measures on financial performance.

The bank had employed a bonus scheme for the employees in its retail branches, based on quarterly financial targets. If the target was met, a branch manager had to allocate a bonus pool among the employees in the branch. Prior to the intervention, branch managers had no access to information on the sales made by individual employees. Hence, managers distributed individual bonuses based on subjective performance assessments. From July 2003 until December 2003, managers

---

[23] Note that this is of course a subjective rating in itself and not an objective measure of performance.

in a treatment group of 23 branches gained access to objective sales performance measures for each of their employees across different product categories; this was announced two months before the intervention. Nothing was changed in the control group of the remaining (more than 250) branches, and the rules of the bonus system remained also otherwise unchanged.

The analysis of the experiment reveals a causal effect of having objective performance measures on branch performance. The intervention increased the number of employee-initiated customer appointments by 11 % after it was announced. It raised profits by about 2 % on average and 5 % in the largest branches, even though the intervention came at no cost to the bank. Interestingly, the intervention had no effect in smaller branches, which is in line with the idea that it is easier for supervisors in small branches to keep track of employee performance, even when no objective performance information is available. In larger branches, however, this is more difficult, and here the accessibility of objective performance information had a significantly larger effect on incentives.[24] Hence, the field experiment shows that providing objective performance information can indeed be beneficial, in particular when it is hard for supervisors to keep track of all employees.

4.5 *Differentiation and Employee Perceptions: Descriptive Field Evidence*

But of course, in many jobs it is simply infeasible or prohibitively costly to access objective performance indicators. In these settings, firms need to rely on subjective evaluations. The question then remains whether firms should foster differentiation. As we have seen above, there are trade-offs involved. Differentiation seems to help when employees work separately, but it may be detrimental when employees can easily harm each other without being observed. Indeed, when Yahoo recently introduced a forced distribution, many articles in the press complained about this change (example headlines are "Forced Ranking Is as Bad for Yahoo as it Was for Microsoft" (Forbes), "Yahoo's Latest HR Disaster: Ranking Workers on a Curve" (businessweek.com), and "Yahoo is Ranking Employees: When Microsoft Did that, it Was a Disaster" (washingtonpost.com)). Hence, an interesting question is whether employee satisfaction is indeed lower in firms that foster differentiation.

We study this question with data from the first wave of the Linked Personnel Panel (LPP) described in section 3, where we observe whether a firm employs a recommended distribution for performance evaluations or not (we do not observe whether firms use a forced distribution, which is very rare in Germany). As this is purely cross-sectional data, we caution that we cannot identify a causal effect by using such an instrument here. However, we can answer the question whether,

---

[24] A detailed analysis of the performance effects also reveals substantial differences between product categories that also explain the branch-size effects to some extent. There is less separation of labor in smaller branches, which causes multitasking problems. Products where performance was not well measured before the intervention benefited also from a shift in efforts from the core product (consumer loans), where performance actually decreased in smaller branches. The effects are robust, for instance, when individual branches are taken out of the sample or size cutoffs are varied.

ceteris paribus, the use of this practice is a credible signal of lower employee satisfaction.

In the following, we analyze the relationship between the use of recommended distributions in performance evaluations and employee perceptions, such as job satisfaction and fairness preferences, by making use of the matched employer–employee character of the LPP. We estimate individual-level regressions, with different employee perceptions and attitudes as dependent variables. The first item, job satisfaction, is measured by the question "How satisfied are you with your job?" on an 11-point Likert scale. Affective commitment is measured with the six-item short form introduced by Meyer, Allen, and Smith (1993). Work engagement is operationalized with the nine-item short scale of the Utrecht Work Engagement Scale (Schaufeli and Bakker, 2004). Helping and cooperation are reflected by two items measuring how often an employee offers help to her coworkers and how often coworkers themselves offer help in case it is needed. Finally, fairness of compensation is measured by a single item that reflects whether the employee perceives her compensation in the establishment as fair.

The main independent variable is a firm-level dummy variable indicating whether the employing establishment uses recommended distributions in their performance appraisal process. We control for establishment size, monthly net salary (in thousands of euros), gender, part-time work, type of job (white-collar or blue-collar, supervisory position or not), contract type (short- or long-term), age, highest educational attainment, and highest professional qualification, as well as industry and region fixed effects. Robust standard errors clustered on the establishment level are reported in all regressions.

The table below shows that recommended distributions do not come along with reduced job satisfaction, commitment, and work engagement. The coefficients are even positive, but statistically insignificant. Surprisingly, employees report significantly higher levels of helping behavior and perceive their compensation to be fairer in firms that use recommended distributions, as columns 4 and 5 reveal. Again it is important to note that this should not be interpreted causally. We believe that the most plausible interpretation for this finding is that firms that are better managed have more professional appraisal systems in place.[25] Guidelines about the distribution of grades are an element of many professional appraisal systems, because otherwise, for instance, different supervisors would follow different standards in the same firm. The causal links can be further explored when longitudinal data on employee perceptions becomes available. But we can now already conclude that the use of recommended distributions is not a negative, but if anything a positive, signal about the perceived quality of work in a firm.

---

[25] Further regressions additionally controlling for proxies of better management such as the existence of variable payment schemes, personnel development plans, written target agreements, employee feedback talks, and workforce planning show a reduced coefficient for recommended distributions, which supports the argument that the use of recommended distributions is rather a signal of good management and the coefficients are not estimates of a causal effect.

Table
Recommended Distribution and Employee Outcomes

	(1) Job satis- faction	(2) Commit- ment	(3) Work en- gagement	(4) Helping	(5) Fair com- pensation
Recommended	0.0551	0.0569	0.0098	0.0606**	0.1208**
distribution	(0.0787)	(0.0425)	(0.0343)	(0.0275)	(0.0563)
Monthly	0.2301***	0.1486***	0.0408**	0.0117	0.279***
net salary	(0.0419)	(0.0260)	(0.0207)	(0.0148)	(0.0527)
Constant	6.774***	3.041***	3.481***	4.268***	2.781***
	(0.1731)	(0.0898)	(0.0763)	(0.0610)	(0.1407)
Observations	3,627	3,586	3,521	3,617	3,623
R^2	0.038	0.109	0.053	0.015	0.124

Notes: Additional control variables: dummies for female, part-time, white-collar, short-term contract, management position, age, highest educational attainment, highest professional qualification, establishment size, industry, and region. Robust standard errors clustered on establishment level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5 Discussion

We argue that it is important to apply the toolbox of different research methods when studying management practices, also in order to help firms to design better practices. Formal economic models help to develop a precise understanding of potential behavioral mechanisms. Laboratory experiments allow for the isolation and disentanglement of these mechanisms in precisely controlled environments. Field experiments in firms help us to estimate the causal effects of instrument use on the performance of firms. And, finally, the use of broad representative surveys among firms and employees gives us more detailed information about the frequency and correlates of its use in real companies and allows us to study the generalizability of the insights gained. Hence, it is important to stress the complementary character of these different approaches.

In our view, the use of these complementary methods is particularly necessary when studying performance appraisals, a core HR practice in most firms; for the behavioral and economic mechanisms involved can be surprisingly intricate and complex. But the presented research also reveals some robust patterns that imply rules of thumb for the design of appraisal systems: Differentiation increases performance when the interdependence between the assessed employees is not too strong, but it may increase incentives for counterproductive behavior, especially when cooperation and teamwork are important. Too fine-grained differentiation without objective performance information may do harm by violating reference points of employees, and negative reciprocal reactions may then outweigh potential positive incentive effects. Objective performance measurement can help by avoiding rating biases. However, if objective performance information is not available

and firms have to rely on subjective assessments, they should try to *manage* these assessments. Otherwise each supervisor is guided by her own individual social preferences, which leads to inconsistent evaluation standards across different units of a firm. This may well explain why employees in firms with recommended distributions are not unhappier and even perceive a higher fairness of compensation.

To conclude, we strongly believe that academic research using the presented mix of complementary methods can inform the practitioner's debate and help to gain a broader understanding of what drives individual behavior in firms.

References

Abbink, K., and A. Sadrieh (2009), "The Pleasure of Being Nasty," *Economics Letters*, 105(3), 306–308.

Baker G., M. Gibbs, and B. Holmstrom (1994a), "The Internal Economics of the Firm: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 109(4), 881–919.

—, —, and — (1994b), "The Wage Policy of the Firm," *The Quarterly Journal of Economics*, 109(4), 921–955.

Bandiera, O., I. Barankay, and I. Rasul (2011), "Field Experiments with Firms," *The Journal of Economic Perspectives*, 25(3), 63–82.

Bartel, A., C. Ichniowski, and K. Shaw (2004), "Using 'Insider Econometrics' to Study Productivity," *The American Economic Review*, Papers and Proceedings, 94(2), 217–223.

Bell, D. E. (1985), "Disappointment in Decision Making under Uncertainty," *Operations Research*, 33(1), 1–27.

Bellemare, C., and B. Shearer (2009), "Gift Giving and Worker Productivity: Evidence from a Firm-Level Experiment," *Games and Economic Behavior*, 67(1), 233–244.

Bellmann, L., S. Bender, M. Bossler, S. Broszeit, C. Dickmann, et al. (2015), "LPP – Linked Personnel Panel: Quality of Work and Economic Success: Longitudinal Study in German Establishments (data collection on the first wave)," FDZ-Methodenreport 5/2015, The Research Data Centre (FDZ) of the Federal Employment Agency in the Institute for Employment Research, Nuremberg.

Berger, J., C. Harbring, and D. Sliwka (2013), "Performance Appraisals and the Impact of Forced Distribution—An Experimental Investigation," *Management Science*, 59(1), 54–68.

Bernardin, H. J., D. K. Cooke, and P. Villanova (2000), "Conscientiousness and Agreeableness as Predictors of Rating Leniency," *Journal of Applied Psychology*, 85(2), 232–236.

Black, S. E., and L. M. Lynch (2004), "What's Driving the New Economy?: The Benefits of Workplace Innovation," *The Economic Journal*, 114(493), F97–F116.

Bloom, N., C. Genakos, R. Sadun, and J. Van Reenen (2012), "Management Practices across Firms and Countries," *Academy of Management Perspectives*, 26(1), 12–33.

—, B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013), "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 128(1), 1–51.

— and J. Van Reenen (2007), "Measuring and Explaining Management Practices across Firms and Countries," *The Quarterly Journal of Economics*, 122(4), 1351–1408.

— and — (2010), "Why do Management Practices Differ across Firms and Countries?" *The Journal of Economic Perspectives*, 24(1), 203–224.

— and — (2011), "Human Resource Management and Productivity," in: O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics: Volume 4B*, North-Holland, Amsterdam, pp. 1697–1767.

Bol, J. C. (2011), "The Determinants and Performance Effects of Managers' Performance Evaluation Biases," *The Accounting Review*, 86(5), 1549–1575.

Bolton, G. E. (1991), "A Comparative Model of Bargaining: Theory and Evidence," *The American Economic Review*, 81(5), 1096–1136.

— and A. Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition," *The American Economic Review*, 90(1), 166–193.

Breuer, K., P. Nieken, and D. Sliwka (2013), "Social Ties and Subjective Performance Evaluations: An Empirical Investigation," *Review of Managerial Science*, 7(2), 141–157.

Camerer, C. F. (2015), "The Promise and Success of Lab–Field Generalizability in Experimental Economics: A Reply to Levitt and List," in: G. R. Fréchette and A. Schotter (eds.), *Handbook of Experimental Economic Methodology*, Oxford University Press, Oxford, pp. 249–295.

Combs, J., Y. Liu, A. Hall, and D. Ketchen (2006), "How Much do High Performance Work Practices Matter? A Meta-Analysis of their Effects on Organizational Performance," *Personnel Psychology*, 59(3), 501–528.

Deaton, A. (2010), "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 48(2), 424–455.

Delfgaauw, J., R. Dur, J. Sol, and W. Verbeke (2013), "Tournament Incentives in the Field: Gender Differences in the Workplace," *Journal of Labor Economics*, 31(2), 305–326.

Dohmen, T. J., B. Kriechel, and G. A. Pfann (2004), "Monkey Bars and Ladders: The Importance of Lateral and Vertical Job Mobility in Internal Labor Market Careers," *Journal of Population Economics*, 17(2), 193–228.

Dur, R., and J. Tichem (2015), "Altruism and Relational Incentives in the Workplace," *Journal of Economics & Management Strategy*, 24(3), 485–500.

Engellandt, A., and R. T. Riphahn (2011), "Evidence on Incentive Effects of Subjective Performance Evaluations," *Industrial & Labor Relations Review*, 64(2), 241–257.

Englmaier, F., A. Roider, and U. Sunde (2012), "The Role of Salience in Performance Schemes: Evidence from a Field Experiment," Discussion Paper 6448, IZA, Bonn.

— and K. Schüßler (2016), "Complementarities of HRM Practices: A Case for a Behavioral Economics Perspective," *Journal of Institutional and Theoretical Economics (JITE)*, 172(2), forthcoming.

Falk, A., and J. J. Heckman (2009), "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science*, 326(5952), 535–538.

Fehr, E., S. Gächter, and G. Kirchsteiger (1997), "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica*, 65(4), 833–860.

—, G. Kirchsteiger, and A. Riedl (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108(2), 437–459.

— and K. M. Schmidt (1999), "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114(3), 817–868.

— and — (2006), "The Economics of Fairness, Reciprocity, and Altruism – Experimental Evidence and New Theories," in: S. Kolm and J. M. Ythier (eds.), *Handbook of the Economics of Giving, Altruism, and Reciprocity*, Volume 1, North-Holland, Amsterdam, pp. 615–691.

Flabbi, L., and A. Ichino (2001), "Productivity, Seniority and Wages: New Evidence from Personnel Data," *Labour Economics*, 8(3), 359–387.

Fletcher, C., and R. Williams (1996), "Performance Management, Job Satisfaction and Organizational Commitment," *British Journal of Management (BJM)*, 7(2), 169–179.

Frederiksen, A., F. Lange, and B. Kriechel (2012), "Subjective Performance Evaluations and Employee Careers," Discussion Paper 6373, IZA, Bonn.

Friebel, G., M. Heinz, M. Krüger, and N. Zubanov (2015), "Team Incentives and Performance: Evidence from a Retail Chain," Discussion Paper 9316, IZA, Bonn.

Gibbs, M., K. A. Merchant, W. A. Van der Stede, and M. E. Vargus (2004), "Determinants and Effects of Subjectivity in Incentives," *The Accounting Review*, 79(2), 409–436.

Giebe, T., and O. Gürtler (2012), "Optimal Contracts for Lenient Supervisors," *Journal of Economic Behavior & Organization*, 81(2), 403–420.

Grund, C., and J. Przemeck (2012), "Subjective Performance Appraisal and Inequality Aversion," *Applied Economics*, 44(17), 2149–2155.

— and D. Sliwka (2009), "The Anatomy of Performance Appraisals in Germany," *The International Journal of Human Resource Management*, 20(10), 2049–2065.

Halse, N., V. Smeets, and F. Warzynski (2011), "International Differences in Subjective Performance Evaluation, Compensation and Career Dynamics in a Global Company," Working Paper 11–15, Department of Economics, University of Aarhus.

Heckman, J. J. (1992), "Randomization and Social Program Evaluation," in: C. F. Manski and I. Garfinkel (eds.), *Evaluating Welfare and Training Programs*, Harvard University Press, Cambridge (MA), pp. 201–230.

Hossain, T., and J. A. List (2012), "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations," *Management Science*, 58(12), 2151–2167.

Huselid, M. A. (1995), "The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance," *The Academy of Management Journal*, 38(3), 635–672.

— and B. E. Becker (1996), "Methodological Issues in Cross-Sectional and Panel Estimates of the Human Resource-Firm Performance Link," *Industrial Relations*, 35(3), 400–422.

Imbens, G. W. (2010), "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48(2), 399–423.

Kahneman, D., and A. Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263–291.

Kampkötter, P. (2016), "Performance Appraisals and Job Satisfaction," *The International Journal of Human Resource Management*, Published Online First January 1, DOI: 10.1080/09585192.2015.1109538.

—, J. Mohrenweiser, D. Sliwka, S. Steffes, and S. Wolter (2016), "Measuring the Use of Human Resources Practices and Employee Attitudes: The Linked Personnel Panel," *Evidence-based HRM: A Global Forum for Empirical Scholarship*, forthcoming.

— and D. Sliwka (2014), "The Role of Differentiation in Performance Evaluations – Theory and Evidence," unpublished Manuscript, University of Cologne.

Kane, J. S., H. J. Bernardin, P. Villanova, and J. Peyrefitte (1995), "Stability of Rater Leniency: Three Studies," *The Academy of Management Journal*, 38(4), 1036–1051.

Kőszegi, B., and M. Rabin (2006), "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 121(4), 1133–1165.

Landy, F. J., and J. L. Farr (1980), "Performance Rating," *Psychological Bulletin*, 87(1), 72–107.

— and — (1983), *The Measurement of Work Performance: Methods, Theory, and Applications*, Academic Press, New York.

Levitt, S. D., and J. A. List (2007), "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *The Journal of Economic Perspectives*, 21(2), 153–174.

Loomes, G., and R. Sugden (1986), "Disappointment and Dynamic Consistency in Choice under Uncertainty," *The Review of Economic Studies*, 53(2), 271–282.

Ludwig, J., J. R. Kling, and S. Mullainathan (2011), "Mechanism Experiments and Policy Evaluations," *The Journal of Economic Perspectives*, 25(3), 17–38.

Manthei, K., and D. Sliwka (2014), "Multitasking and the Benefits of Objective Performance Measurement – Evidence from a Field Experiment," unpublished Manuscript, University of Cologne.

Meyer, J. P., N. J. Allen, and C. A. Smith (1993), "Commitment to Organizations and Occupations: Extension and Test of a Three-Component Conceptualization," *Journal of Applied Psychology*, 78(4), 538–551.

Michaels, E., H. Handfield-Jones, and B. Axelrod (2001), *The War for Talent*, Harvard Business Press, Boston (MA).

Moers, F. (2005), "Discretion and Bias in Performance Evaluation: The Impact of Diversity and Subjectivity," *Accounting, Organizations and Society*, 30(1), 67–80.

Murphy, K. J. (1992), "Performance Measurement and Appraisal: Motivating Managers to Identify and Reward Performance," in: W. J. Burns, Jr. (ed.), *Performance Measurement, Evaluation, and Incentives*, Harvard Business School Press, Boston (MA), pp. 37–62.

Murphy, K. R., and J. N. Cleveland (1991), *Performance Appraisal: An Organizational Perspective*, Allyn & Bacon, Boston (MA).

— and — (1995), *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives*, SAGE Publications, Thousand Oaks (CA).

Ockenfels, A., D. Sliwka, and P. Werner (2015), "Bonus Payments and Reference Point Violations," *Management Science*, 61(7), 1496–1513.

Prendergast, C. (1999), "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37(1), 7–63.

— and R. Topel (1993), "Discretion and Bias in Performance Evaluation," *European Economic Review*, 37(2–3), 355–365.

— and R. H. Topel (1996), "Favoritism in Organizations," *Journal of Political Economy*, 104(5), 958–978.

Rynes, S. L., B. Gerhart, and L. Parks (2005), "Personnel Psychology: Performance Evaluation and Pay for Performance," *Annual Review of Psychology*, 56, 571–600.

Schaufeli, W., and A. Bakker (2004), *Utrecht Work Engagement Scale*, Preliminary Manual, Version 1.1, Occupational Health Psychology Unit, Utrecht University.

Sebald, A., and M. Walzl (2014), "Subjective Performance Evaluations and Reciprocity in Principal–Agent Relations," *The Scandinavian Journal of Economics*, 116(2), 570–590.

Subramony, M. (2009), "A Meta-Analytic Investigation of the Relationship between HRM Bundles and Firm Performance," *Human Resource Management*, 48(5), 745–768.

Takahashi, S., H. Owan, T. Tsuru, and K. Uehara (2014), "Multitasking Incentives and Biases in Subjective Performance Evaluation," Discussion Paper Series A 614, Institute of Economic Research, Hitotsubashi University.

Varma, A., A. S. Denisi, and L. H. Peters (1996), "Interpersonal Affect and Performance Appraisal: A Field Study," *Personnel Psychology*, 49(2), 341–360.

Welch, J., and J. A. Byrne (2001), *Jack: Straight from the Gut*, Time Warner Book Group, New York.

Whitman, D. S., D. L. Van Rooy, and C. Viswesvaran (2010), "Satisfaction, Citizenship Behaviors, and Performance in Work Units: A Meta-Analysis of Collective Construct Relations," *Personnel Psychology*, 63(1), 41–81.

Zizzo, D. J., and A. J. Oswald (2001), "Are People Willing to Pay to Reduce Others' Incomes?" *Annales d'Économie et de Statistique*, (63/64), 39–65.

Patrick Kampkötter
Seminar of Personnel Economics and
Human Resource Management
University of Cologne
Albertus-Magnus-Platz
50923 Köln
Germany
patrick.kampkoetter@uni-koeln.de

Dirk Sliwka
Seminar of Personnel Economics and
Human Resource Management
University of Cologne
Albertus-Magnus-Platz
50923 Köln
Germany
dirk.sliwka@uni-koeln.de

New Frontiers in Empirical Research
on Informal Contracting

by

Ricard Gil and Giorgio Zanarone[*]

In standard models of informal contracting, parties use discretionary payments to split the known value of their relationship, which results in stable efficient contracts over time. Relaxing simplifying assumptions, recent models shed light on how informal contracts evolve over time, the creation of relational rents, and their sensitivity to path dependence. This paper investigates how these nonstandard theories of informal contracts can be tested empirically. We first discuss predictions from a selection of representative models, and strategies necessary to test them. We then examine how existing evidence supports these predictions, and how available data may be used for further testing. (JEL: D23, L24, M21)

1 Introduction

At least since Macaulay (1963), economists, sociologists, and legal scholars alike have documented the pervasiveness of *informal contracts* – that is, contracts enforced by the threat of terminating valuable relationships, rather than by courts. Building on this fundamental insight, economists have developed a standard principal–agent framework for analyzing informal contracts, which we refer to as the *standard model* (e.g., MacLeod and Malcomson, 1989; Baker, Gibbons, and Murphy, 1994, 2002; Levin, 2003; Malcomson, 2013). This framework has generated a number of predictions on the determinants of informal contracting, and on how informal and formal contracts interact with one another, spinning a small but growing body of empirical evidence (see Gil and Zanarone, 2015).[1]

[1] In a companion paper, Gil and Zanarone (2015) review empirical strategies and opportunities in the existing empirical literature that tests implications of the standard informal contracting model. Our analysis here extends and complements that paper in that

While undoubtedly valuable, the standard model is based on some strong simplifying assumptions, among which particularly important are:

(1) symmetric information about the value of relationships, and
(2) unconstrained discretionary payments that can be used to share the long-term rents from cooperation.

Because of these assumptions, the standard model predicts that optimal informal contracts are stationary (i.e., they do not change over time, conditional on the state of nature), and thus it cannot explain how contractual relationships are built and gradually evolve, and how their evolution may be subject to path dependence (e.g., Gibbons and Henderson, 2012, 2013; Helper and Henderson, 2014).

An important part of the new theoretical frontier on informal contracts is to relax simplifying assumptions (1) and (2) in the standard model, and develop *nonstandard* theories that can explain the patterns discussed above. For instance, MacLeod (2003), (the last part of) Levin (2003), and Fuchs (2007) analyze models of subjective performance evaluation where the agent's output is privately observed by the principal, providing important insights into the use of efficiency wage contracts backed by the threat of termination, and on the timing of performance reviews by the principal. Chassang (2010) sheds light on routine building and path dependence in informal relationships, studying a model where parties in an informal contract have private information on each other's payoff functions, and hence on the optimal definition of cooperation. Similarly, Halac (2012), which is discussed in greater detail below, explores the consequences of asymmetric information (in her case, on the parties' outside options) for the dynamic evolution of performance bonuses in relational incentive contracts. Finally, Li and Matouschek (2013) show how asymmetric information on an employer's ability to pay his employee due to contingent cash constraints results in employment relationships characterized by permanent cycles of high and low cooperation.

Regarding the unlimited-liability assumption in the standard model, a number of recent papers have relaxed it in a diverse number of settings such as the choice between loyal and new partners in supply networks (Board, 2011, discussed below), the dynamic patterns of knowledge transfers in apprenticeships (Garicano and Rayo, 2013), the cycles of cooperation within firms (Fong and Li, 2015), the distribution of new technologies in developing economies (Fuchs, Green, and Levine, 2015), and the path dependence of informal delegation of authority in organizations (Li, Matouschek, and Powell, 2014).[2]

As mentioned above and also discussed in our companion paper (Gil and Zanarone, 2015), a body of empirical evidence on the standard model of informal contracting is slowly emerging. However, there is no evidence to date on the nonstandard models described above. In this paper, we aim to encourage the develop-

---

we discuss the empirical testability of theoretical predictions that require modifying or removing some of the assumptions in the standard informal contracting model.

[2] See Ray (2002) for a more general model of informal contracting with limited liability and for references to earlier studies.

ment of such an empirical research frontier, by providing a methodological framework for testing the emerging nonstandard models of informal contracting.

The article is organized as follows. Section 2 summarizes the theoretical predictions from selected models of informal contracting that relax assumptions (1) and (2) in the standard model. Section 3 discusses the data necessary to test those predictions, as well as potential empirical strategies. Section 4 discusses the testability of other recent models of informal contracting that further depart from the standard model, and section 5 concludes.

2 Theoretical Predictions

Because the theoretical literature is still at its outset, there is no unified nonstandard model from which to derive general testable predictions. Hence, we will follow a heuristic approach – that is, we will discuss two representative nonstandard models (Board, 2011; Halac, 2012) and how to test them, hoping that interested researchers will use our framework as a starting point to push the frontier further. We choose these papers over others not because we believe they are qualitatively superior, but because, besides being published in top journals in economics, they seem especially well fitted for empirical exploration. In particular, and as discussed in greater detail below, both papers offer relatively simple comparative-static predictions that should impose manageable data collection burdens on empirical researchers. Therefore, they seem a reasonable starting point to discuss the testability of nonstandard informal-contracting models. In the final section of the paper, we will briefly discuss the testability of nonstandard models other than Halac (2012) and Board (2011).

2.1 Informal Contracts under Incomplete Information: The Halac Model

Halac (2012) relaxes assumption (1) above in the standard model – namely, symmetric information – and studies informal incentive contracts in settings where the agent does not know how much the principal values her future relationship with him. Let principals who do and do not value the relationship be called *high-commitment* and *low-commitment* types, respectively. Compared to those of low-commitment types, the profits of high-commitment principals are more heavily reduced if their relationship with the agent terminates – for instance, because the agent possesses human capital that is specific to them, or because they have less time to search for replacements. Hence, upon observing satisfactory agent's performance, high-commitment principals can credibly promise to pay him a larger informal incentive bonus, because they know the agent can more severely punish them by threatening to terminate the relationship.

A key result in Halac (2012) is that in general, the agent cannot infer whether the principal is a high-commitment or low-commitment type from the incentive contract the principal negotiates with her at the beginning of the relationship. In

particular, and in stark contrast to the standard case of complete information, the agent cannot infer that a principal who offers a large incentive bonus values the relationship, and hence can credibly commit to pay. To understand the intuition behind this result, suppose the agent is offered a large informal bonus and, based on that, she believes the principal to be a high-commitment type. Since the agent trusts the principal to pay, she responds to the bonus scheme by exerting high effort. Then, a low-commitment principal would have a strong incentive to offer the same high-powered incentive contract, as that would allow him to gain from the agent's high effort, and then cut her off-guard by reneging on the promised bonus.

Building on this insight, Halac (2012) goes on to show that both the initial incentive contract and its evolution in the course of the relationship importantly change depending on whether the principal or the agent has bargaining power. We elaborate on this point below.

2.1.1   Case 1: The Principal Has Bargaining Power

This will be the case, for instance, if the agent does not belong to a union. Since the principal cannot signal his commitment level ex ante through the size of the offered bonus, the agent can only detect low-commitment principals through ex post signaling – that is, when the principal reneges on the promised bonus payments. As a consequence, all principals offer the same incentive contract, and the baseline bonus payment (that is, the first-period bonus) is high-powered if the agent holds a strong prior belief that the principal is a high-commitment type, in which case the risk of default is limited, whereas it is low-powered otherwise.

Regarding the relationship's dynamics, if the agent believes the principal to have high commitment, the optimal incentive contract is designed in such a way that the promised bonus payments and the agent's effort, but also the probability of default by the principal, increase over time, as the agent grows more and more convinced of the principal's commitment. In contrast, when the agent believes the principal to have low commitment, the optimal incentive contract is stationary, and it is characterized by low bonus payments, low effort, and no defaults, in every period.

2.1.2   Case 2: The Agent Has Bargaining Power

This case may be relevant, for instance, if the agent belongs to a (strong) union, or if she has political support. Halac (2012) shows that, in this case, the agent can infer the principal's commitment level ex ante, through screening. Under screening, the agent demands a high-powered incentive contract, characterized by identically large payments and effort requirements in each period, and designed in such a way that if the principal honors the contract, his discounted stream of profits equals his outside option. While a high-commitment principal accepts (and honors) the proposed contract, a low-commitment principal prefers to reject it, because he knows that if he accepted the agent's offer he would subsequently renege on the

promised payment, so the relationship would prematurely terminate, and hence his expected stream of profits would not compensate the outside option. Once the low-commitment principal reveals himself by rejecting the proposed contract, there is one period of *strike,* without production, after which the agent negotiates with the principal a low-powered incentive contract, compatible with his limited commitment ability, from the next period onward.

Halac (2012) also shows that, since screening is potentially costly (due to the period of strike), it is only optimal if the agent's prior belief that the principal has high commitment, and hence will accept the contract, is strong. If that is not the case, the agent prefers to offer a low-powered incentive contract that both principal types are willing to accept.

2.1.3 Summary of Testable Predictions

From the results discussed above, a number of potentially testable predictions can be obtained. First, at the beginning of a contractual relationship, bonuses in an informal incentive contract should be higher, relative to those in an optimal formal contract, when the agent has a strong prior belief that the principal values the relationship, and hence is able to commit.

Second, the likelihood of defaults on informal bonuses by the principal should increase in the agent's prior belief when the principal has bargaining power, but not when the agent has bargaining power, as in the latter case ex ante screening ensures that there are no defaults in equilibrium.

Finally, informal bonus payments, and the probability of default, should be increasing over time when the principal (the informed party) has bargaining power, and nonincreasing over time when the agent has bargaining power. This last prediction depends on the fact that when the agent has bargaining power, the probability of default is zero. Moreover, bonus payments do not change over time if the agent's prior assessment of the principal's commitment is low, or if the agent's prior assessment is high and it turns out to be correct ex post. Finally, if the agent's prior assessment is high but the principal turns out to have low commitment, bonus payments are high in the first period, and drop to a low level from the second period onwards.

2.2 *Informal Contracts under Liquidity Constraints: The Board Model*

Board (2011) relaxes assumption (2) in the standard model – namely, the absence of liquidity constraints – and studies informal contracts between a principal and multiple agents.

In his baseline application, the principal is a buyer, and the agents are suppliers who may hold up the buyer's specific investments. The suppliers' actions and the relationships' outcomes are nonverifiable, so holdup can only be prevented by resorting to informal contracts sustained by repeated interactions. In any given period, the buyer observes the cost of investing in each available supplier, after which

he decides on which supplier to invest in. Board (2011) assumes the suppliers are cash-constrained, so the buyer cannot make them residual claimants while extracting the relationship's value via upfront fees. As a result, optimal informal contracts must grant long-term quasi-rents to the suppliers, so that when a given supplier is used, he will not be tempted to hold up the buyer for fear of losing those rents.

When starting his business, the buyer informally announces his choice of suppliers for all future periods, contingent on the realized costs. Absent liquidity constraints, the buyer would extract the full surplus from each supply relationship, so he would allocate suppliers efficiently by investing in the lowest-cost supplier in each period. Since suppliers are cash-constrained, though, the buyer reckons that every time he switches to a new supplier he needs to pay him a stream of quasi-rents to prevent holdup. As a result, the buyer is biased against new suppliers, in the sense that he may prefer to rely on the existing suppliers, or *insiders*, even in periods where new suppliers, or *outsiders*, are less costly, in order to save on quasi-rents. Board (2011) calls this inefficient reliance on insider suppliers *loyalty*.

2.2.1 Summary of Testable Predictions

As discussed by Board (2011), the broad prediction that optimal contracts exhibit loyalty to insiders is consistent not only with his informal contracting model, but also with models where there are exogenous costs of switching suppliers – for instance, training of the new supplier's employees. However, in an exogenous-switching-cost model with complete information, the degree of loyalty should not depend on whether the contract between buyer and suppliers is formal or informal (if the buyer's information on suppliers were incomplete, this might no longer be true, because a "bad" outside supplier would have more opportunities to renege on an informal contract than on a formal one, and so he would be more risky than an insider for the buyer). Moreover, in an exogenous-switching-cost model where loyalty is not part of an informal agreement between suppliers and the buyer, a supplier's performance should not depend on whether the buyer is loyal to her (in the sense that he assigns her more business than to outsiders).

Based on these observations, a number of potentially testable predictions can be obtained from Board's model.

First, a buyer's loyalty to its suppliers should be greater in countries characterized by lower court quality, where contracts tend to be informal.

Second, if a buyer stops being loyal to his current suppliers and switches to outsiders, the insiders should decrease the quality of their performance in future interactions with the buyer.

Third, buyers who decide to become disloyal and switch to new suppliers should use more formal contracts in their future dealings with inside suppliers, as they can no longer rely on informal agreements to prevent holdup.

Finally, a buyer's loyalty to his suppliers should be greater when the expected duration of the relationship is short – that is, when the common discount factor of buyer and suppliers is small. The reason for this last prediction is that, as the

discount factor decreases, the stream of quasi-rents that the buyer must grant his suppliers to prevent holdup increases, and so does the cost of switching to outside suppliers.

3   Testability

3.1   Potential Empirical Tests of Halac (2012)

In order to test the predictions from Halac (2012), one needs to measure at least four variables: (i) the incentive bonus offered at the beginning of an informal principal–agent relationship; (ii) the evolution of informally contracted bonus payments in the course of the relationship; (iii) the allocation of bargaining power between the principal (a company, a manager, a buyer) and the agent (a CEO, a worker, a seller); and (iv) the agent's prior beliefs regarding the principal's type (low- or high-commitment type in the model). Once measures for these variables are available, one can test the predictions of Halac's model on how information and bargaining power affect the terms and dynamic evolution of informal incentive contracts. Although Halac (2012, p. 772) argues that her analysis is applicable to a broad range of scenarios such as "employment contracts, inter-firm agreements, supply-chain relationships, informal credit contracts, and other settings where contracting tends to be informal and information is typically incomplete," we will primarily focus here on a few well-known data sets on employment and procurement contracts. At the end of this section, we will also propose a new empirical setting – namely, incentive contracts in academia – that seems to fit the mechanisms described in Halac (2012), and hence may be used to test the predictions from her model.

3.1.1   Data

Since assessing the empirical validity of the predictions from Halac (2012) requires following the principal–agent relationship over time, we may need longitudinal data on informal incentive payments starting from the beginning of a contractual relationship, as well as across relationships that vary in their distribution of bargaining power and degree of informational asymmetry. Collecting data on whether incentive payments are formal or informal is challenging but possible, as demonstrated by existing studies of employment contracts (Gillan, Hartzell, and Parrino, 2009) and interfirm distribution agreements (Gil, 2013). The agent's beliefs about the type of the principal can be assessed by looking at whether the agent has some knowledge of the principal's history and reputation (perhaps through public rankings such as Fortune's "Best Companies to Work for" index) or has specific expertise to judge the principal's type (for instance, because he has been an employee or frequent business partner of the principal in the past). Finally, variables that measure the bargaining power of the agent would be the degree of unionization of labor force in a company or location, and whether the principal is a monopsonist (e.g., sole employer for all workers) in the local labor market.

If data on payments are available but it is not possible to establish whether they are informal, an indirect test could be developed by looking at whether the patterns predicted by Halac (2012) are stronger in settings where courts are weak, and hence contracts (even if formal) can be presumed to be informally enforced. Measures for the availability of court enforcement across countries are proposed, among others, by Johnson, McMillan, and Woodruff (2002) and by Antràs and Foley (2014).

3.1.2 Empirical Strategies

A test of Halac's (2012) predictions on the effect of the agent's information and bargaining power on the initial size and dynamics of informal incentives could be performed in the context of executive compensation by combining the empirical strategies of Gillan, Hartzell, and Parrino (2009) and DeVaro, Kim, and Vikander (2015). Gillan, Hartzell, and Parrino (2009) study the choice between explicit and implicit employment agreements (EAs) for CEOs in S&P 500 firms. The interesting feature of their data is that it exploits a property of the legal environment – namely, the obligation to disclose all formal dimensions of CEO employment agreements to the SEC – to measure whether the compensation terms in such agreements are formal (if disclosed) or informal (if not disclosed). Aside from these data, DeVaro, Kim, and Vikander (2015) have the complementary, longitudinal information on the size of CEOs' bonuses year by year, which is necessary to test Halac's predictions on the dynamics of informal incentives. As suggested by Gillan, Hartzell, and Parrino (2009), the beliefs of the agent about the principal's type can be measured in this context by looking at whether the CEO has been promoted from inside the company or has been hired from another company in the same industry. An outside CEO should be less informed about the company's past bonus practices, and hence, controlling for observable characteristics of the company, should be more uncertain about the company's trustworthiness. Moreover, the bargaining power of CEOs can be measured by proxies for their outside options, such as the number of other companies in the same industry.

In a procurement context, Ahmadjian and Oxley (2006, 2013) examine the relationships between buyers and input suppliers in Japanese conglomerates (*keiretsu*). In these networks, buyers and suppliers usually own each other's stocks and participate in each other's governance through membership in the board of directors and shareholder meetings. Whether stock cross-ownership exists (and the share of buyer's stock owned by the supplier) may be a good proxy for the supplier's (agent's) prior belief that the buyer (principal) values the informal relationship with its partner. Their data set also specifies how dependent each supplier is on the buyer's purchases, and vice versa, which may be a good proxy for bargaining power. Similar measurements could be obtained from the European survey data used by Johnson, McMillan, and Woodruff (2002) and, more recently, Calzolari et al. (2014). The two major drawbacks in those data (relative to Ahmadjian and Oxley's papers) are that stock cross-ownership is less common in Europe than in Japan, and that survey data may suffer from the subjectivity of the responses.

Notice that none of the aforementioned procurement data sets has information on the size and evolution of incentive payments and on whether those payments are informal, so these data sets need to be augmented in order to test the predictions of Halac (2012).

In traditional franchising networks, such as car dealerships, companies can unilaterally determine the maximum sales margins earned by their distributors by changing the wholesale prices and discounts. For instance, as discussed in Zanarone (2009, 2013), dealership contracts give European car manufacturers full discretion to change the list price of cars, which is the maximum that can be charged to customers, and the discounts awarded to dealers. The higher the expected sales markups in year $t + 1$, the stronger a dealer's incentive to invest in innovative marketing campaigns, store maintenance, and customer comfort in year t. From this point of view, sales margins can be seen as informal incentive payments. One may then test the predictions from Halac (2012) by collecting longitudinal data on the maximum sales margins, which are typically recorded in annexes to the dealership contract. The dealers' prior beliefs may be approximated by measures of the manufacturer's reputation, such as its position in "best franchise" rankings or the number of years it has been franchising (Arruñada, Garicano, and Vázquez, 2001). Bargaining power may be approximated by the manufacturer's market share, by whether the dealer owns multiple stores, or by whether dealers are collectively organized through an association.

Finally, although there are no existing data sets to be exploited, a context where in our opinion Halac (2012) could be tested is academia, particularly economics departments and business schools (our main sources of anecdotal knowledge). In academia, deans and department chairs make decisions in a rather discretionary way on bonuses and salary raises awarded to faculty members, contingent on good research and teaching performance, on adjustments to teaching loads and schedules in response to unforeseen events, on tenure decisions, and on future recruitment, positioning, and growth strategies. A new department or school, or one that has decided to change its focus from teaching to research, will advertise itself in the Job Openings for Economists and attend international job market venues such as the American Economic Association (AEA) or Academy of Management meetings. There, the new department will face a lineup of potential candidates who may wonder whether the department can credibly commit to its promises on discretionary compensation, teaching loads and schedules, tenure requirements, and consistency in recruiting efforts.

Then, a potential way to test Halac's (2012) predictions on the relationship between informal incentives and bargaining power in academia is to compare the offers received by candidates from fields that are in higher and in lower demand (for instance, finance versus economics). Similarly, one may test Halac's predictions on how the agent's prior belief about the principal's trustworthiness affects incentives by comparing the offers made by newly created departments from universities with different reputations (for instance, as measured by official rankings), where reputations would proxy for the candidates' belief that their prospective em-

ployer is a trustworthy type. A potential problem in conducting this study is data availability. One possible strategy would be to run a large-scale survey. Since most of the information discussed above is fairly objective, response biases would be less of a concern than in other types of surveys. Information on which departments or schools began participating in the international academic job market in any given year may be obtained from the archives of the AEA and of corresponding societies in fields other than economics.

3.1.3 Potentially Related Evidence

As we saw, Halac's (2012) model predicts that incentive bonuses and effort may increase over time when the principal has private information on his own commitment ability. When that is the case, the joint surplus – that is, the expected value of the informal relationship between the principal and the agent – should also increase over time, as the relationship unfolds and the agent updates his belief about the principal's commitment.

In their study of contractual relationships between Kenyan flower exporters and their international clients, Macchiavello and Morjaria (2015) provide evidence consistent with this prediction. First, they estimate (lower bounds of) the value of these contractual relationships by computing the revenues forgone by flower exporters when delivering a batch to a client at the stipulated price rather than selling it on the spot market. Because contracts are in fact informal due to the low quality of Kenyan courts, self-enforcement requires that the value of the relationship be at least as large as the exporter's temptation to sell on the spot market. Second, Macchiavello and Morjaria (2015) show that the estimated lower bound of the contractual relationship for a given seller–client pair increases over time as the relationship grows older.

While the evidence in Macchiavello and Morjaria (2015) is consistent with Halac's (2012) prediction, their model differs from Halac (2012) in that it assumes symmetric uncertainty and learning about the seller's reliability, rather than asymmetric information, and as a result, it does not predict signaling, screening, or breach in equilibrium. It is therefore unclear whether the predictions unique to Halac (2012), on the relationship between contractual dynamics, prior beliefs at the beginning of the relationship, and the parties' relative bargaining power, could be tested using the data set of Macchiavello and Morjaria (2015).

We will return later on in this section to the potential empirical relevance of Macchiavello and Morjaria's (2015) data set for testing nonstandard models of informal contracting.

3.2 Potential Data Sets to Test Board (2011)

Ideally, to test Board's (2011) main predictions on a principal's propensity to switch agents, one needs longitudinal data on the identity of a firm's contractual partners, and on whether these partners have been contracting with the firm before.

As Board (2011) points out, the ideal settings for testing this model are nonexclusive supply agreements (e.g., Asanuma, 1989). To separate Board's theory from theories that yield similar predictions but do not involve informal contracts (for instance, those appealing to the differential skills of new agents versus insiders), one would also need to identify whether the contractual relationship is informal or not. As before, this could be done indirectly, by comparing institutional settings with strong and with weak court enforcement.

There are several data sets used for other purposes that could shed light on the validity of Board's (2011) predictions. Here we discuss three of them in some detail.

First, Gil and Marion (2013) use data on highway procurement auctions in California where contractors must choose a number of subcontractors when they submit their bids. Their analysis documents the existence of loyalty: all else equal, contractors are more likely to choose subcontractors with which they have worked before. Their study does not examine loyalty literally as defined by Board (2011) – that is, as a contractor's preference for inefficient insider subcontractors over efficient but outsider ones. A variable that has been used in the literature to measure efficiency is distance of the subcontractor from the project location. Therefore, a more precise measure for loyalty here may be a contractor's propensity to choose a subcontractor with which it has interacted before even when it is not the one located closest to the project. Gil and Marion (2013) also provide an exogenous inverse measure for a contractor's and a subcontractor's intertemporal discount rate, which they proxy by the number and value of projects auctioned over the next calendar year.[3] Hence, the data in Gil and Marion (2013) may be used to test the fourth testable prediction from Board's model – namely, on the positive relationship between loyalty and the discount rate.[4]

Second, the aforementioned study of Gil and Marion (2013), as well as the study of flower import–export in Kenya by Macchiavello and Morjaria (2015), may be used to test Board's predictions that the quality of suppliers' services should decrease after the buyer stops being loyal and brings in outside suppliers. This prediction could be indirectly tested using the data in Gil and Marion (2013). While those data do not include measures of the quality of subcontractors' services, they do provide longitudinal information on a contractor's profits and sales, which should be positively affected by the subcontractors' quality. Board's prediction may be more directly tested using the data in Macchiavello and Morjaria (2015), which include

---

[3] Since projects are offered when they are needed and not as a response to a particularly good match between contractor and subcontractor, this measure of the potential future value of the relationship is not contaminated by endogenous considerations of how well contractors and subcontractors are matched or were matched in the past.

[4] Calzolari et al. (2014) use a data set of relationships between German carmakers and their input suppliers. Their data come from a cross-sectional survey and may contain questions that bear on relationship strength and loyalty. Their data may also be used to test whether carmakers in Germany are less likely to switch to new suppliers when these are located in countries in which the quality of the legal system differs.

longitudinal information on the matching between flower exporters and importers, as well as on whether an exporter delivers a specific shipment of roses to its client on time. To test Board's prediction, one may look at whether an exporter's deliveries are less likely to be fulfilled on time after his client has introduced new exporters into his supply chain.

Finally, in order to mitigate the holdup problem, such disloyal firms will try to introduce more complete contracts. Moszoro, Spiller, and Stolorz (2014) use a large data set of publicly available contracts from the SEC's EDGAR database, constructing variables that proxy contract length and rigidity. We may test this fourth implication by comparing firms that change partners often with those that are likely to stay with the same partner. This prediction would imply that the contracts of these two types of firms look very different in length and rigidity, and more so for projects where holdup is more likely to occur.

3.3 Control Variables and Caveats

While the theoretical models in both Halac (2012) and Board (2011) deliver sharp predictions on informal contracting under asymmetric information and liquidity constraints, some of those predictions may be also delivered by models of formal contracts or by more conventional models of informal contracts. In this section, we briefly discuss what empirical controls would be needed to distinguish Board's and Halac's predictions from those of competing models.

A first caveat, discussed in section 2.2, pertains to Board's (2011) prediction that the degree of loyalty should decrease in the extent of formal contracting (for instance, as measured by court quality). This prediction is unique to Board's model under the assumption that information is complete irrespective of whether suppliers are insiders or outsiders. If the buyer's information on outside suppliers (for instance, on their financial situation or honesty) were instead incomplete, the buyer might be more concerned about contracting with outsiders than with insiders – that is, loyalty might arise in equilibrium. Moreover, the buyer's loyalty might increase as formal contracts become harder to enforce, because absent formal contracts, the buyer would have fewer safeguards against the outside suppliers' default or dishonest behavior.

Hence, the tests of Board's prediction that we suggest above would be stronger and more reliable if one could control for the buyer's information on suppliers. As we suggested in discussing the testability of Halac's model, public information on a company's reliability may be provided by official rankings and other reputational measures, such as credit ratings, as well as by ownership cross-holdings and membership in a business association (Ahmadjian and Oxley, 2006, 2013; Johnson, McMillan, and Woodruff, 2002).

Finally, most of Board's (2011) predictions might no longer be unique if there were a systematic difference between the supplies a buyer procures from insiders and outsiders. For instance, suppose a buyer procures one product from supplier *A* exclusively (say, because *A* has unique skills), whereas he is considering switch-

ing to a cheaper outside supplier *B* for a second product. If the contracts between the buyer and his suppliers are informal, the buyer may choose to be loyal to *A* because if he reneges on *A* he stands to lose more, as *A* supplies him with two products, whereas *B* would only supply him with one product. In other words, enforcing multiproduct informal contracts may be easier, as suggested by Bernheim and Whinston (1990) and Levin (2002). Hence, empirical tests of Board's model would be stronger if one could control for the scope of a buyer's contractual relationships with insiders, relative to outsiders.

4   Testing Other Nonstandard Models of Informal Contracting

In this last section, we briefly describe two additional nonstandard informal contracting models, and we discuss their potential testability. As we discuss below, these models offer predictions on important dimensions of firm organization, such as cooperation cycles and informal delegation, which are not studied in Halac (2012) and Board (2011). At the same time, these models seem to impose a tougher data collection burden on empirical researchers. That is why our discussion of testability here is mostly tentative.

4.1   Li and Matouschek (2013)

Like Halac (2012), Li and Matouschek (2013) modify the informational structure of the standard informal contracting model. The difference is that, while Halac (2012) assumes the principal has private information on his type, Li and Matouschek (2013) assume the principal has private information on his cost of paying the agent. Their model is thus well suited to analyze conflicts that arise when the principal fails to pay a bonus despite observing high effort, and the agent cannot tell whether this is due to the principal's unforeseen financial problems or to reneging (see Berg and Fast, 1975, and Stewart, 1993, for anecdotal evidence that these conflicts occur in employment relationships).

If the agent is to be motivated by the promise of an informal bonus, the principal's failure to pay such a bonus must be punished, so, unlike in the standard model, surplus destruction occurs in equilibrium in Li and Matouschek (2013). More specifically, they show that under an optimal informal contract: (1) the relationship is characterized by indefinitely alternating cycles of cooperation and punishment; (2) punishment cycles are "smooth," in the sense that effort and surplus decrease gradually after the principal's privately observed shock, whereas cooperation cycles are "steep," in the sense that effort and surplus revert to the efficient level as soon as the principal starts honoring bonus payments again; and finally, (3) the relationship does not terminate in equilibrium.[5]

---

[5] Other models of informal agreements that predict cycles of cooperation and punishment are Green and Porter (1984); Yared (2010); Englmaier and Segal (2014); and Contreras (2015).

To (indirectly) test predictions (1) and (2) in Li and Matouschek (2013), one would need data on effort or output in an employment relationship, or in a supply or distribution relationship, over a long time horizon.[6] In addition, one would need to control for product and business cycles in order to fully support Li and Matouschek's (2013) informal-contracting explanation for cyclic performance. There are a few studies that use effort-level data in employment relationships (see Mas and Moretti, 2009, on Safeway cashiers; and Bandiera, Barankay, and Rasul, 2005, and related papers, on fruit-picking). Several measures of output have been proposed in buyer–supplier relationships and procurement, such as trade credit levels (Johnson, McMillan, and Woodruff, 2002), delays and cancellations in air transportation (Forbes and Lederman, 2009), contractors' ability to bid low in auctions (Gil and Marion, 2013), and exporters' timely deliveries to importers (Macchiavello and Morjaria, 2015).

A more direct test could be performed by running a survey of contractual relationships that are at least in part informal and collecting data on how deviations are punished. As discussed before, one may be able to classify a contractual relationship as informal by relying on company records (Gil, 2013) or by looking at contracts that are subject to mandatory disclosure (Gillan, Hartzell, and Parrino, 2009) or occur in countries with weak court enforcement (Johnson, McMillan, and Woodruff, 2002; Antràs and Foley, 2014).

4.2 Li, Matouschek, and Powell (2014)

Li, Matouschek, and Powell (2014) use an informal contracting model where the agent has private information about a decision-relevant state to study power dynamics in organizations – that is, how power is earned, lost, and retained by individuals or groups within organizations. The authors note that informal incentives in organizations are often provided by a promise of empowerment rather than by money, so it is important to understand how informal power dynamics are designed to make informal incentive contracts self-enforcing. As in Green and Porter (1984) and Li and Matouschek (2013), optimal informal contracts entail equilibrium punishments in Li, Matouschek, and Powell (2014). Unlike in those models, though, power delegation (empowerment) does not cycle back and forth between principal and agent, and ends either in permanent centralization or in permanent empowerment. Sadly enough, whether the organization ends in one or the other steady state is fully determined by random events that occur during the organization's history.

Li, Matouschek, and Powell (2014) also find that under an informal empowerment contract, and in contrast with learning-based models, an organization's ability to adapt to changes in the environment gets *worse* over time: as a result of past empowerment promises, the agent may end up being fully and permanently em-

---

[6] As Li and Matouschek (2013) recognize, prediction (2) on the slope of cycles is sensitive to their assumption that the production function is concave.

powered, and thus he may cease to choose projects that are in the best interest of the principal.

Their final prediction is that these power dynamics in organizations may partly explain the observed large and persistent differences across seemingly similar enterprises (Gibbons and Henderson, 2013), and why mature firms may have a harder time adapting to changes in their environments than younger rivals. These differences in performance may arise across industries and firms even when copying strategies and organizational practices of successful rivals is possible.

Corporate governance is a context where the predictions from Li, Matouschek, and Powell (2014) may be tested. By now, there is evidence that entrenched (staggered) boards are associated with a reduction in firm value (Bebchuk and Cohen, 2005). The question is whether, as predicted by Li, Matouschek, and Powell (2014), these companies are also less likely to adapt to changes in the industry. To test for that, we would need to empirically observe whether, besides being less valuable, companies with entrenched boards are also associated with especially large losses in value when an industry shakeout occurs or upon entry of new competitors.

Finally, what makes Li, Matouschek, and Powell (2014) a difficult paper to test is that its underlying mechanisms rely heavily on factors that are hard to observe. First, the number and type of projects available to managers are usually unobservable to the econometrician (in the same way that they are unobservable to the principal in their model), and therefore, it is difficult to verify when, why, and how a manager may be abusing her power by choosing a project that is not in the best interest of the shareholders. Second, power within organizations is difficult to quantify. Hence, an empirical study, testing whether the mechanisms underneath the negative relationship between board entrenchment and firm performance conform to the model in Li, Matouschek, and Powell (2014), should follow firms and managers over time and quantify the concentration of power through hierarchical control decision authority and the like. While this may be possible (e.g., Guadalupe, Li and Wulf, 2014), it does impose a heavy data collection burden on the empirical researcher.

5   *Conclusion*

In this paper we have discussed the testability of recent theoretical models of informal contracting, which study relationships characterized by incomplete information and imperfect ability to share rents. In such environments, informal contracts change over time in ways that respond to the parties' attempts to maximize rent extraction (Board, 2011), resolve informational asymmetries (Halac, 2012), and punish imperfectly monitored deviations (Li and Matouschek, 2013).

While a body of empirical evidence on these nonstandard models does not exist yet, testing them seems important, as they have implications for fundamental managerial choices like the dynamic structure of informal incentives in employment (Halac, 2012; Li and Matouschek, 2013) and the degree of turnover in supply

chains (Board, 2011). To contribute to filling this gap between theory and evidence, we have suggested possible tests of the nonstandard models, which could be performed using existing data on employment relationships as well as on supply and distribution chains. We hope our paper will foster future empirical research that will expand our understanding of this new frontier in informal contracting theory, and help us to push the frontier even further.

References

Ahmadjian, C., and J. Oxley (2006), "Using Hostages to Support Exchange: Dependence Balancing and Partial Equity Stakes in Japanese Automotive Supply Relationships," *The Journal of Law, Economics, & Organization*, 22(1), 213–233.

Ahmadjian, C. L., and J. E. Oxley (2013), "Vertical Relationships, Hostages, and Supplier Performance: Evidence from the Japanese Automotive Industry," *The Journal of Law, Economics, & Organization*, 29(3), 485–512.

Antràs, P., and C. F. Foley (2014), "Poultry in Motion: A Study of International Trade Finance Practices," *Journal of Political Economy*, 123(4), 853–901.

Arruñada, B., L. Garicano, and L. Vázquez (2001), "Contractual Allocation of Decision Rights and Incentives: The Case of Automobile Distribution," *The Journal of Law, Economics, & Organization*, 17(1), 257–284.

Asanuma, B. (1989), "Manufacturer–Supplier Relationships in Japan and the Concept of Relation-Specific Skill," *Journal of the Japanese and International Economies*, 3(1), 1–30.

Baker, G., R. Gibbons, and K. J. Murphy (1994), "Subjective Performance Measures in Optimal Incentive Contracts," *The Quarterly Journal of Economics*, 109(4), 1125–1156.

—, —, and — (2002), "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 117(1), 39–84.

Bandiera, O., I. Barankay, and I. Rasul (2005), "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 120(3), 917–962.

Bebchuk, L. A., and A. Cohen (2005), "The Costs of Entrenched Boards," *Journal of Financial Economics*, 78(2), 409–433.

Berg, N. A., and N. D. Fast (1975), "The Lincoln Electric Company," Case Study 376-028, Harvard Business School, Boston (MA).

Bernheim, B. D., and M. D. Whinston (1990), "Multimarket Contact and Collusive Behavior," *The RAND Journal of Economics*, 21(1), 1–26.

Board, S. (2011), "Relational Contracts and the Value of Loyalty," *The American Economic Review*, 101(7), 3349–3367.

Calzolari, G., L. Felli, J. Koenen, G. Spagnolo, K. O. Stahl (2014), "Trust, Competition and Innovation: Theory and Evidence from German Car Manufacturers," Working Paper 5229, Center for Economic Studies & ifo Institut (CESifo), Munich.

Chassang, S. (2010), "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts," *The American Economic Review*, 100(1), 448–465.

Contreras, O. (2015), "Relational Contracts, Financing Constraints and Social Welfare," http://extranet.sioe.org/uploads/isnie2014/contreras.pdf, last accessed on January 24, 2016.

DeVaro, J., J.-H. Kim, and N. Vikander (2015), "Pay-for-(Persistent)-Luck: CEO Compensation under Relational Contracting," unpublished Manuscript, California State University, East Bay.

Englmaier, F., and C. Segal (2014), "Cooperation and Conflicts in Organizations," unpublished Manuscript, Ludwig-Maximilians-Universität München.

Fong, Y-f., and J. Li (2015), "Relational Contracts, Limited Liability, and Employment Dynamics," https://www.kellogg.northwestern.edu/faculty/li/htm/efficiency%20Wage%20150511.pdf, last accessed on January 24, 2016.

Forbes, S. F., and M. Lederman (2009), "Adaptation and Vertical Integration in the Airline Industry," *The American Economic Review*, 99(5), 1831–1849.

Fuchs, W. (2007), "Contracting with Repeated Moral Hazard and Private Evaluations," *The American Economic Review*, 97(4), 1432–1448.

—, B. Green, and D. I. Levine (2015), "Optimal Arrangements for Distribution in Developing Markets," unpublished Manuscript, Haas School of Business, University of California, Berkeley.

Garicano, L., and L. Rayo (2013), "Relational Knowledge Transfers," Discussion Paper 1203, Centre for Economic Performance (CEP), London School of Economics and Political Science.

Gibbons, R., and R. Henderson (2012), "Relational Contracts and Organizational Capabilities," *Organization Science*, 23(5), 1350–1364.

— and — (2013), "What do Managers Do?: Exploring Persistent Performance Differences among Seemingly Similar Enterprises," in: Gibbons and Roberts (eds.) (2013, pp. 680–731).

— and J. Roberts (eds.) (2013), *The Handbook of Organizational Economics*, Princeton University Press, Princeton (NJ).

Gil, R. (2013), "The Interplay of Formal and Relational Contracts: Evidence from Movies," *The Journal of Law, Economics, & Organization*, 29(3), 681–710.

— and J. Marion (2013), "Self-Enforcing Agreements and Relational Contracts: Evidence from California Highway Procurement," *The Journal of Law, Economics, & Organization*, 29(2), 239–277.

— and G. Zanarone (2015), "On the Determinants and Consequences of Informal Contracting," available at http://ssrn.com/abstract=2560520.

Gillan, S. L., J. C. Hartzell, and R. Parrino (2009), "Explicit versus Implicit Contracts: Evidence from CEO Employment Agreements," *The Journal of Finance*, 64(4), 1629–1655.

Green, E. J., and R. H. Porter (1984), "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52(1), 87–100.

Guadalupe, M., H. Li, and J. Wulf (2014), "Who Lives in the C-Suite? Organizational Structure and the Division of Labor in Top Management," *Management Science*, 6(4), 824–844.

Halac, M. (2012), "Relational Contracts and the Value of Relationships," *The American Economic Review*, 102(2), 750–779.

Helper, S., and R. Henderson (2014), "Management Practices, Relational Contracts, and the Decline of General Motors," *The Journal of Economic Perspectives*, 28(1), 49–72.

Johnson, S., J. McMillan, and C. Woodruff (2002), "Courts and Relational Contracts," *The Journal of Law, Economics, & Organization*, 18(1), 221–277.

Levin, J. (2002), "Multilateral Contracting and the Employment Relationship," *The Quarterly Journal of Economics*, 117(3), 1075–1103.

— (2003), "Relational Incentive Contracts," *The American Economic Review*, 93(3), 835–857.

Li, J., and N. Matouschek (2013), "Managing Conflicts in Relational Contracts," *The American Economic Review*, 103(6), 2328–2351.

—, —, and M. Powell (2014), "Power Dynamics in Organizations," https://www.kellogg.northwestern.edu/faculty/li/htm/LMP%20Power%20Dynamics.pdf, last accessed on January 24, 2016.

Macaulay, S. (1963), "Non-Contractual Relations in Business: A Preliminary Study," *American Sociological Review*, 28(1), 1–19.

Macchiavello, R., and A. Morjaria (2015), "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports," *The American Economic Review*, 105(9), 2911–2945.

MacLeod, W. B. (2003), "Optimal Contracting with Subjective Evaluation," *The American Economic Review*, 93(1), 216–240.

— and J. M. Malcomson (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57(2), 447–480.

Malcomson, J. M. (2013), "Relational Incentive Contracts," in: Gibbons and Roberts (eds.) (2013, pp. 1014–1065).

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

Moszoro, M. W., P. T. Spiller, and S. Stolorz (2014), "Rigidity of Public Contracts," unpublished Manuscript, University of California at Berkeley-Haas.

Ray, D. (2002), "The Time Structure of Self-Enforcing Agreements," *Econometrica*, 70(2), 547–582.

Stewart, J. (1993), "Taking the Dare," *The New Yorker*, July 26, pp. 34–39.

Yared, P. (2010), "A Dynamic Theory of War and Peace," *Journal of Economic Theory*, 145(5), 1921–1950.

Zanarone, G. (2009), "Vertical Restraints and the Law: Evidence from Automobile Franchising," *The Journal of Law & Economics*, 52(4), 691–700.

— (2013), "Contract Adaptation under Legal Constraints," *The Journal of Law, Economics, & Organization*, 29(4), 799–834.

Ricard Gil   Giorgio Zanarone
Johns Hopkins Carey Business School   Colegio Universitario de
100 International Drive   Estudios Financieros (CUNEF)
Baltimore, MD 21202   C/ Leonardo Prieto Castro 2
U.S.A.   28040 Madrid
ricard.gil@jhu.edu   Spain
                                         gzanarone@cunef.edu