HARVARD | BUSINESS | SCHOOL

What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises

Robert Gibbons
Rebecca Henderson

Working Paper

13-020

August 17, 2012

What Do Managers Do?
Exploring Persistent Performance Differences among Seemingly Similar Enterprises*

Robert Gibbons

Rebecca Henderson

23:30, July 23, 2012

* We are very grateful for help from Dan Barron, Nick Bloom, Sylvain Chassang, Eliza Forsythe, Marina Halac, Casey Ichniowski, Nico Lacetera, Ruitian Lang, Hongyi Li, Jin Li, Maria Polyakova, Mike Powell, Jan Rivkin, John Roberts, Annalisa Scognamiglio, Kathryn Shaw, Chad Syverson, John Van Reenen, Sarah Venables, and seminar audiences at Brown, Columbia, Dartmouth, Emory, Harvard, LBS, LSE, University of Michigan, University of Minnesota, Stanford, Sydney, Tilburg University, Unviersity of Toronto, Vanderbilt, Washington University, and Yale. This project would not have been possible without years of crucial support from MIT Sloan's Program on Innovation in Markets and Organizations, especially Nancy Beaulieu, Nelson Repenning, John Sterman, and Tommy Wang.

What Do Managers Do?

Exploring Persistent Performance Differences among Seemingly Similar Enterprises

Robert Gibbons and Rebecca Henderson

1. Introduction

Decades of research using a wide variety of detailed plant- and firm-level data has provided strong evidence of persistent performance differences among seemingly similar enterprises (hereafter, PPDs among SSEs). Bartelsman and Doms (2000) reviewed the sizeable initial literature on this issue, and a recent review by Syverson (2011) highlighting much new research has only strengthened the result.

As one striking example, Syverson (2004a) finds that "within 4-digit SIC [Standard Industrial Classification] industries in the U.S. manufacturing sector, the average difference in logged total factor productivity (TFP) between an industry's 90th and 10th percentile plants is 0.651 … [meaning that] the plant at the 90th percentile of the productivity distribution makes almost *twice* as much output *with the same measured inputs* as the 10th percentile plant." And the U.S. is not exceptional: Hsieh and Klenow (2009) "find even larger productivity differences in China and India, with average 90-10 TFP ratios over 5:1" (Syverson 2011: 326-327, emphasis in the original).

Indeed, the existence of persistent performance differences among seemingly similar enterprises is so well established that recent work in trade, industry dynamics, and productivity growth has increasingly taken this fact as a fundamental starting point. For example Melitz's (2003: 1695) celebrated paper begins: "Recent empirical research using longitudinal plant or firm-level data from several countries has overwhelmingly substantiated the existence of large and persistent productivity differences among establishments in the same narrowly defined industries."

The question of what *causes* these performance differences, however, is still wide open. Syverson (2011) offered a long list of possibilities, including differences in management practice, higher quality labor and capital, differential investment in information technology and research and development (R&D), learning by doing, firm structure, productivity spillovers, regulatory behavior, and differences in competitive regime. Here we focus on the first of these: one important source of PPDs among SSEs may be persistent differences in management practices.

We choose this focus for four reasons, which we state briefly here and unpack below. First, in large-sample studies, differences in management practices are correlated with differences in productivity. Second, focused-sample studies at the plant and even the line level suggest that these large-sample results are robust to controls for many other factors that Syverson identifies as potential determinants of productivity. Third, choosing management practices as a key driver of PPDs surfaces an important question rarely asked in economics but central to strategic management: why do best practices not diffuse more readily? Finally, we hope that our specific discussion of management practices and PPDs will help shape a broader research agenda in organizational economics about what, exactly, managers do.

We begin from the observation that many competitively significant management practices cannot be reduced to well-defined action rules that can be specified ex ante and verified ex post. Instead, the implementation of these management practices is critically dependent on context. For example, for many years the singularly successful retailer Nordstrom asked its sales associates to "use their good judgment in all situations." Similarly, Toyota's chief request of its production workers was that they seek to "continuously improve the production process." The state-dependent actions necessary to meet these expectations cannot be the subject of a formal contract. We therefore focus on relational contracts—roughly, understandings that the parties share about their roles in and rewards from cooperating together, but understandings so rooted in the details of the parties' relationship that they cannot be shared with a court.

Viewed broadly, we see our perspective on management practices as consistent with decades of work on management and authority, such as Barnard (1938), Simon (1947), Penrose (1959), Cyert and March (1963), and Arrow (1974). As one crisp summary, Mintzberg (2004) separates a manager's activities into analysis (deciding what to do) and administration (getting

the organization to do it). Although both are surely important, the strong theme we see in the literature from Barnard to Arrow focuses on the latter, as do we here. In short, we see relational contracts as a key way that managers get organizations to get things done. Furthermore, we see many competitively significant management practices as relying on relational contracts that themselves are hard to build and change, leading to the slow diffusion of management practices that could improve organizational performance.

The core of this chapter is therefore (1) evidence that some important management practices rely on relational contracts, (2) evidence that relational contracts can be hard to build and change, and (3) examples of recent theories that begin to address some of these issues. To prepare for these core aspects of the chapter, in Sections 2 and 3 we briefly review the two literatures mentioned above that motivate our argument. In Section 2 we summarize evidence that PPDs among SSEs exist and are economically significant across a wide range of industries and geographies. These results have been shown to be robust to concerns that they reflect problems of selection, or simultaneity, or the distinction between productivity in revenue versus physical terms. These productivity differences have also been shown to be persistent over time and surprisingly resistant to increased competition. That is, as we describe in Section 2, an increase in competition does tend to increase an industry's average productivity and decrease its productivity dispersion, but such dispersion is by no means eliminated, even in very competitive environments.

In Section 3 we then summarize evidence that performance differences are importantly correlated with variation in management practices. We complement the recent large-sample work in economics by drawing particularly on focused-sample research in industrial relations and human resource management that has explored the role of high-commitment work systems in driving productivity, as well as focused-sample research in strategy and organizational studies that has studied the role of white-collar work practices in explaining competitive advantage. As we describe, such focused-sample work can offer especially sharp measures of control variables, dependent variables, and the independent variables of interest (i.e., management practices); see Ichniowski and Shaw (this volume) for more on such "insider econometrics."

In the remainder of the chapter we consider relational contracts. We begin Section 4 by sketching the basic theory of relational contracts, which is closely related to more established

work on repeated games. We then describe qualitative examples of competitively significant management practices that rely on relational contracts—at Lincoln Electric and Toyota for blue-collar work and at Merck for white-collar work.

Section 5 turns to the question of why management practices associated with persistent performance differences do not diffuse more readily. We begin by reviewing the extensive literature on this topic, which has developed along four lines. First, managers may have problems of *perception*—they do not know they are behind. Second, managers may have problems of *inspiration*—they know they're behind, but they don't know what to do about it. Third, managers may have problems of *motivation*—they know they're behind and they know what to do, but they lack incentive to adopt new practices. Fourth, managers may have problems of *implementation*—they know they're behind, they know what to do, and they're trying hard to do it, but they nonetheless cannot get the organization to get it done.[1]

Given the evidence from Section 4 that important management practices may depend on relational contracts, we continue Section 5 by sketching recent models in which superior performance indeed rests on relational contracts, but these contracts are either too expensive or not even feasible for some enterprises to implement. In short, in these models, bad performance is due to bad parameters among enterprises that are only seemingly similar (i.e., there is unmeasured heterogeneity in the costs of using the key relational contracts).

Our central interest, however, is in the important class of cases where practices do not diffuse even though they are widely acknowledged to be competitively significant, knowledge of how to implement them is reasonably widespread, and firms are striving mightily to adopt them. (Think of the mature phases of the Toyota Production System or Total Quality Management, for example.) Such cases are not well explained either by the perception, inspiration, or motivation arguments for why practices do not diffuse or by relational-contract models in which the relational contracts that underlie high performance are either too expensive or not feasible for some firms. Instead, for the class of cases that is our central interest, we require an implementation argument.

[1] We thank Jan Rivkin for explaining his "four 'tion" labels to us and allowing us to adapt them here for our purposes.

The remainder of the chapter therefore focuses on implementation difficulties that arise through relational contracts. In particular, we suggest two approaches to modeling such issues. We present the first approach in the remainder of Section 5. In very recent relational-contracting models, such as those where path dependence in relational contracting can produce measured performance differences among ex ante identical enterprises, several interesting possibilities arise: cooperation can be hard to build well, in the sense that achieving perfunctory cooperation can make it harder to achieve consummate cooperation; cooperation, once built, can be fragile; and cooperation may be difficult to build in the first place. In all these models, however, the parties play the optimal equilibrium of the game as a whole, and they understand this equilibrium from the beginning of the game. That is, although we have moved beyond explanations of bad performance based on bad parameters, now we have explanations based on bad luck.

Our second approach to modeling implementation difficulties that arise through relational contracts is much more speculative; in fact, we know of no existing models. In Section 6 we therefore draw from both case studies and lab experiments to suggest that the implementation of some relational contracts requires solving problems of not only *credibility* but also *clarity*. The credibility problem is familiar: should one party believe another's promise? The clarity problem is new: can one party understand another's promise? Furthermore, these problems seem likely to interact: for example, if one party does something unexpected, does the other attribute it to miscommunication or gaming?

We emphasize that the clarity problem is likely to be particularly acute in exactly those settings where relational contracts are needed because formal contracts are imperfect or infeasible: where parties recognize that their roles in and rewards from cooperation are not easy to articulate ex ante or verify ex post. More precisely, the existing relational-contracts literature—including the models described in Section 5 (both those focused on bad parameters and those focused on bad luck)—focuses on the credibility problem that arises when the parties' roles in and rewards from cooperation cannot be verified ex post. The additional issue we raise in Section 6 is the clarity problem that arises when these roles and rewards cannot be articulated ex ante. We conclude Section 6 with some highly speculative thoughts about how this clarity problem might be modeled. As our discussion makes clear, we view this section as a very early report on a research agenda that is promising but just beginning.

Of course, much of our argument echoes earlier work, especially outside economics. For example, there are enormous literatures on organizational culture and its potential connection to organizational performance; see Schein (1985) and Barney (1986), respectively. Even closer to our themes, Ostrom (1990) analyzed self-organized and self-governed institutions for managing common-pool resources, Rousseau (1995) studied "psychological contracts" within organizations, and Adler and Heckscher (2006) saw "the firm as a collaborative community." Finally, inside economics, Leibenstein (1969, 1987) suggested that underperforming enterprises (those inside the production possibility frontier, or "X-inefficient") might be stuck in Defect-Defect equilibria, whereas superior performers might have learned to play Cooperate-Cooperate. Not surprisingly, these contributions from outside economics made little use of formal models. Furthermore, although Leibenstein's argument appealed to multiple equilibria familiar from repeated-game models, neither his work nor subsequent models spoke directly to "stuck in" or "learned to." Kreps (1990, 1996) pointed toward formal models of the latter issues, but the literature then went quiet.

To conclude this introduction, we return to our rationales for writing this chapter, which are encapsulated by the two parts of our title. Working backward, we begin with the existence of persistent performance differences among seemingly similar enterprises. As Syverson (2011: 326) notes, "the magnitudes involved are striking," from which it is a short step to Lucas's (1988: 5) remarks about analogous differences in income and growth across countries: "I do not see how one can look at figures like these without seeing them as representing *possibilities*. … This is what we need a theory … *for*: to provide some kind of framework for organizing facts like these, for judging which represent opportunities and which necessities" (emphasis in the original). In short, one needn't aspire to move a 10^{th} percentile firm to the 90^{th}; moving up a quartile would be a big deal, so we need to know whether (and, if so, how) it can be done.

Continuing backward, we reach managers. In particular, we believe that a better understanding of the role of management practices in PPDs among SSEs could have significant policy implications. Since at least Cyert and March (1963), students of organizations have struggled to understand how organizations make decisions; see Gibbons, et al. (this volume) for more. To put it mildly, firms do not always appear to costlessly and constantly optimize their choices from a fixed and known production possibility set, and yet the economic analyses

underlying regulation and other policies often assume this to be so. We expect there to be useful analogies between how a firm struggles to improve its productivity, as considered in this chapter, and how a firm responds to regulations and incentives created by policies: in both cases, not only analysis but also implementation will be involved, and this could change the way we think about policy regarding research funding, patent law, trade liberalization, antitrust rules, and beyond. Both for these policy reasons and more generally, we hope this chapter facilitates future empirical and theoretical work on what managers actually do.

2. PPDs among SSEs Exist and Are Economically Significant

The evidence that PPDs among SSEs exist and are economically significant has been ably reviewed by Bartelsman and Doms (2000) and Syverson (2011). They describe the extensive empirical literature in the area, discussing in some depth the econometric and data-quality issues inherent in accurately measuring heterogeneity in productivity across firms and plants.

This body of research establishes that there exist very significant productivity differences across plants in a wide range of industries and geographies. Furthermore, these findings on productivity differences have been shown to be robust to concerns about selection and simultaneity, as well as the distinction between physical- and revenue-based productivity. They have also been shown to be persistent over time and to be surprisingly resistant to the pressure of increasing competition.

In this section we highlight some of the central papers in the literature. Given these existing surveys, our intention is to be illustrative, not exhaustive. We complement the economics literature with some discussion of research drawn from the strategy and management literatures, including some valuable focused studies.

2.1 Firm-Level Analyses

Early work focused on the relative contribution of "firm" versus "industry" effects in explaining firm-level profitability. Building on the pioneering work of Schmalensee (1985), Rumelt (1991) and McGahan and Porter (1997, 1999) found that business unit effects were significantly more important than industry effects in explaining the dispersion of returns. More recently Hawanini et al (2003) confirm the role of stable firm effects in explaining the variance

in business performance over time and show that this result is robust to multiple measures of value creation. Intriguingly, they show that these results are driven by the performance of firms at the two tails of the distribution: when they drop the top two and the bottom two performers in each industry and re-estimate the models, firm effects fall by 35 - 54% depending on the dependent variable, whereas industry effects increase by nearly 100 - 300%.

Some of the most compelling evidence of persistent firm effects in the data surfaced in the pioneering work of Griliches and Mairesse, who set out to study the effects of R&D on productivity at the firm level. Though not the primary focus of their research, they discovered surprisingly large between-firm heterogeneity in the data (*e.g.*, in deflated sales, number of employees, physical plant, and R&D capital stock) even after accounting for the firm's industrial sector and adjusting for labor inputs. This heterogeneity endured econometric analysis: estimated parameters from a simple model of the production function revealed a large amount of between-firm variability in the firm-specific intercepts and in the slope coefficient for R&D capital (Griliches and Mairesse 1981, 1982, 1985; Griliches 1986).

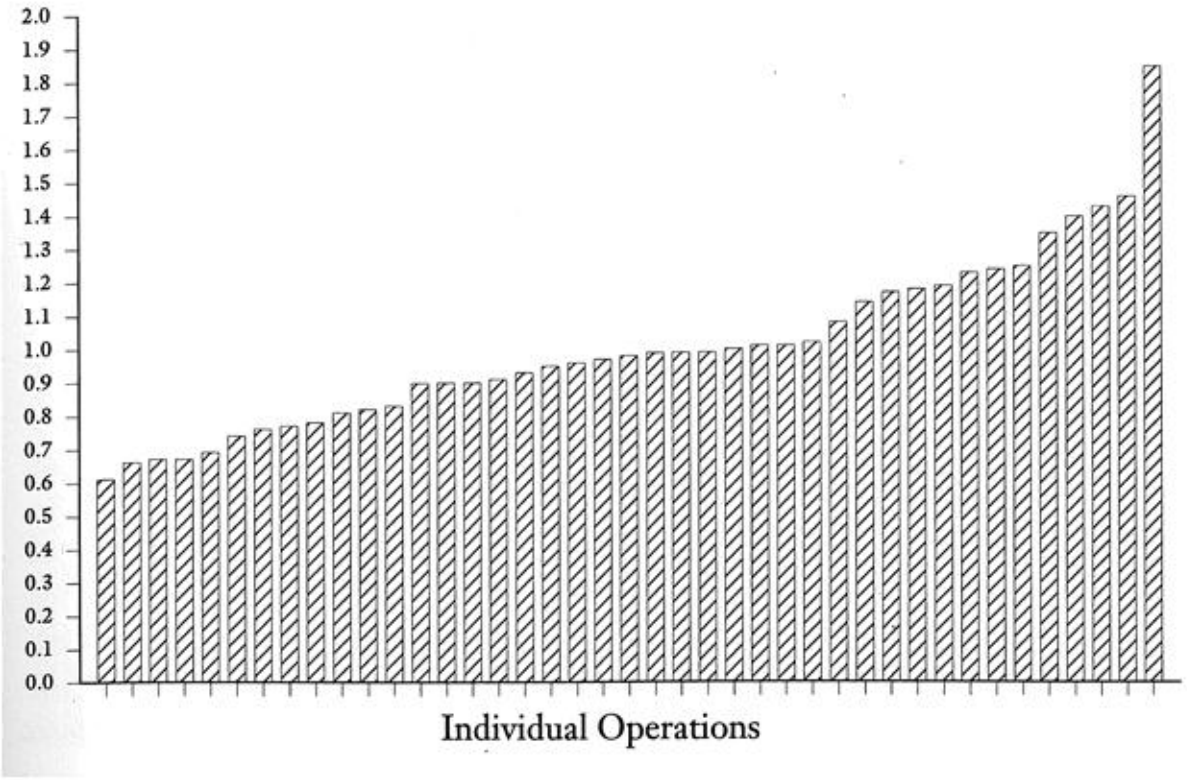
2.2 Plant-Level Analyses

Differences in productivity measured at the firm level immediately raise the question of the degree to which firms are similar. Research at the plant level has addressed this issue by including increasingly sophisticated controls for potential sources of heterogeneity, including product mix, capital vintage, labor quality and market power. We begin by recounting some key accomplishments of the early literature; see Bartelsman and Doms (2000) for more. We follow this with a description of recent contributions considering the distinction between physical and revenue productivity and the econometric challenges inherent in this research.

The exploration of productivity measured at the plant level has a long history, since at least Salter (1960), who studied the pig-iron industry during 1911-1926. Salter found that labor productivity was widely dispersed and that the most productive plant was two times more productive than the average plant. Similarly, Chew et al., (1990) examined productivity dispersion across 41 operating units of a single commercial food division in a large multi-divisional corporation. These 41 operating units all used essentially the same production technology, were all located in the United States, and all produced nearly identical products for

very similar customers. However, even after controlling for such factors as local labor market characteristics, size of the local market, unionization, capital vintage, product quality, and local monopoly power, the top-ranked unit was still twice as productive as the bottom-ranked unit. (Figure 1 shows the raw productivities across units, where the top-to-bottom ratio is 3:1.)

Figure 5-1 Multifactor Productivity Index



Continuing with focused studies, Dunlop and Weil (1996) analyzed 42 business units in the United States that produced a narrow range of apparel products. The raw performance data on lead-time (the total time from ordering inputs to finished products) and operating profits (as a percentage of shipments) exhibited substantial inter-unit variability (coefficients of variation of 0.80 for lead time and 0.69 for operating profits). In a similar spirit, Arthur (1994) reduced product diversity even further in a study of 29 steel mills, but still found substantial coefficients of variation in labor efficiency (0.43) and scrap rates (1.0).

Turning to broader samples, more recent work, using census data from different industries and countries, suggests that the productivity dispersion found in focused-sample studies is

widespread. For example, Dwyer (1998), using panel data from the U.S. census of textile plants in 21 distinct industries, documented 85-15 percentile TFP ratios ranging from 2 to 4 and showed that this dispersion did not decrease over the 1972-1987 time period. Using plant-level data from an even broader set of industries, Syverson (2004a) showed that the average 90-10 percentile labor productivity ratio for 443 four-digit manufacturing industries in 1977 was 7:1; whereas the average 90-10 percentile TFP ratio ranged from 2.7:1 to 1.9:1.

Moving outside the United States, Klette (1999) found significant within-industry differences in permanent plant productivity (as measured by fixed effects) in 13 of 14 Norwegian manufacturing industries studied during 1980-1990. Using similar data, Biorn et al. (2002) estimated a four-factor production function by industry on 4 years of panel data and found that 72-84% of the gross variance in productivity could be attributed to heterogeneity at the plant level. In their analyses of manufacturing establishments in the United Kingdom, Disney et al. (2003) found that establishments in the top decile are 156% more productive than establishments in the bottom decile when productivity is measured according to real gross output per hour of manual labor; this 90-10 productivity advantage was reduced to 91% when log TFP is the productivity measure.

One might be concerned that these productivity results are measured in revenue rather than in physical product. Using a revenue-based measure of TFP, a local monopolist facing limited competition could appear to be more productive than its rival merely because it is able to obtain higher margins for its products. Idiosyncratic demand shifts could also cause high-profitability firms to appear more technically efficient than they really are. However, research controlling for these effects suggests that substantial heterogeneity in productivity remains even when output is measured in physical units. For example, Foster, et al. (2008) use a Census of Manufacturers data set for 1977, 1982, 1987, 1992, and 1997 containing measures of physical output for 11 industries with physically homogeneous goods (e.g. cardboard boxes and white pan bread). The authors document large and persistent within-industry dispersion in physical quantity-based productivity measures. In fact, dispersion in physical quantities is greater than dispersion in revenue-based TFP: after removing product-year fixed effects, the standard deviation for revenue TFP is 22 log points and for physical productivity TFP the standard deviation is 26 log points.

These focused- and broad-sample results also appear to be robust to problems of selection and simultaneity (Olley and Pakes, 1996). For example, Van Biesebroeck (2008) compared five different econometric techniques for estimating productivity. All five techniques produced remarkably similar results regarding productivity heterogeneity in Columbian clothing and apparel industries, with on average one quarter of all firms having TFP less than 45% of median productivity and one quarter of all firms having TFP greater than 42% of median productivity.

2.3 Persistence over Time and under Competitive Pressure

The research we have reviewed thus far has established the existence of cross-sectional productivity differences in many industries and countries. In this section, we review evidence that these productivity differences persist over time. We first review studies based on census data that demonstrate persistence in both the top and the bottom of productivity distributions. We then move on to describe studies that consider the effects of competition in the evolution of productivity distributions; this research suggests that heterogeneous productivity far from disappears as a result of competitive pressure.

In longitudinal data the evidence for the persistence of productivity differentials over time is very strong. In early work, Bailey et al., (1992), computed plant-level TFP for plants in 23 industries at four points in time (1972, 1977, 1982, and 1987), assigning each plant to a quintile of the productivity distribution in each of the survey years. They found substantial persistence in these distributions: 75% of plants in the top quintile in 1972 remained in the top two quintiles in 1977, and 58% plants in these two quintiles remained there in 1982.[2] Persistence was less strong, but still evident, in the bottom of the productivity distribution: 48% of plants in the bottom quintile in 1972 appeared in the bottom two quintiles in 1977, and 54% of plants in the bottom quintile in 1972 remained in the bottom two quintiles in 1982.

Foster et al., (2006) found similar results for firms in retail trade. The authors employed quinquentenial census data (1987,1992, 1997) to compute establishment-level productivity adjusted for four-digit industry effects. They found that "for continuing businesses these differences are highly persistent" (Foster et al. 2006: 757), although exit was also high with the

[2] In the transition matrix, plant productivities are weighted by their employment. With unweighted data, the percentage of 1972 top-quintile plants remaining in the top two quintiles in 1977 is 47%.

entry of national chains. Similarly, Disney et al., (2003) computed the productivity dispersion of manufacturing establishments for a fixed cohort of surviving firms over time. In 1982, the 90-10 spread for the 1982 cohort ranged from 1.48 (ln TFP) to 1.62 (ln labor productivity). Five years later, the spreads for the firms remaining from the 1982 cohort had fallen to 0.81 and 1.43 (ln TFP and ln LP, respectively), but were not close to zero. Even 4 years later, in 1991, the heterogeneity among surviving firms from the 1982 cohort had fallen only a bit further, with 90-10 spreads now 0.66 and 1.28, respectively. Thus, after almost a decade, substantial heterogeneity persisted among those firms that had not exited from an initial cohort of firms.

Turning to the effect of competition, in cross-sectional data, increased competition is associated with increased productivity and reduced productivity dispersion. For example, Syverson (2004b) analyzes cross-sectional data on plants in the ready-mix concrete industry. Very high transportation costs in cement effectively create many distinct geographical markets, and Syverson uses the density of the local construction industry – which varies significantly across markets – as an instrument for local market competition. He finds that, in more competitive markets, average productivity is higher and productivity dispersion is lower (in particular, the productivity distribution is truncated from below). However, even in quite competitive markets, there is substantial dispersion in productivity.

There is also longitudinal evidence that productivity dispersion is highly persistent, even in the face of competitive pressure. Although increasing competition raises average productivity, as weak firms exit and strong firms expand, productivity differences are highly persistent, with some studies even suggesting that greater competitive intensity is associated with a higher degree of plant level heterogeneity. For example, Eslava et al., (2004) use plant-level data from Columbia for 1982-1998 to study the effects of market reforms (trade, labor, and financial) on productivity, profitability, and the reallocation of market share. The authors find TFP dispersion increasing for exiting, entering, and continuing firms in each of the 17 years of their panel, and the persistence in TFP was greater in the years following market reforms. See Holmes et al. (2012) for citations to recent papers with similar findings."

2.4 Focused Studies

Further evidence that productivity differences are real and persistent comes from studies that focus on fewer firms but use much more detailed data, attempting to control at the finest level for differences in capital and labor productivity. For example, the Chew et al., (1990) study of food kitchens was discussed above. As another example, Ichniowski et al., (1997) studied 36 finishing lines in steel minimills. They included more than 25 technological controls in their analysis of the determinants of productivity, including capital vintage, computerization, learning curve effects, line speed and width, the quality of the steel input, equipment design and scheduled downtime for maintenance. They still found very significant productivity differences across lines (which they attribute to differences in human-resource management policies, so we return to this paper in Section 3).

Similarly, in a series of highly detailed field studies of semiconductor manufacturing (Appleyard et al, 2000; Macher and Mowry, 2003, 2009; Macher 2006) Mowery, Macher, and colleagues collected data on defect rates, yield, and cycle time. Consistent with their argument that embedded routines and processes underlie dynamic capabilities, the authors find evidence that interfirm differences in performance are greater than intrafirm differences: "The average within-firm yield coefficient of variation is 1.44, while the between-firm coefficient of variation is 4.42. Similarly, the average within-firm cycle time coefficient of variation is 0.66, while the between-firm coefficient of variation is 2.14. Our performance metrics thus appear to capture firm-specific differences that are both substantial and enduring" (Macher and Mowery, 2009: S51).

In summary, the literature suggests that (1) even after the imposition of tight controls, establishment-level TFP exhibits large dispersion in many industries, countries, and time periods; (2) within-industry productivity rankings are persistent, particularly in the right tail of the productivity distribution; and (3) increased competition leads to aggregate productivity growth but not in a manner that substantially reduces productivity dispersion.

We now turn to the possible role of variation in management practices as one important cause of persistent differences in productivity.

3. PPDs and Variation in Management Practices

Syverson's (2011) review not only documents the existence of PPDs but also categorizes their potential sources into two groups: internal factors (management "levers" in organizations) and external factors (market and environmental factors beyond managerial control). An important next step is to assess the importance of these factors. However, even now a large literature argues that variation in management practices may be an important source of PPDs among SSEs, and it is to this literature that we now turn.

Perhaps the simplest way that managers could be partially responsible for measured performance differences is that some managers are just better than others. Such managerial fixed effects have been discussed at least since Walker (1887) and Mundlak (1961). See Bertrand and Schoar (2003) for recent work blending the two—i.e., a study of how particular managers adopt particular practices. Our focus, however, is on management practices—that is, things that any manager (or, at least, most mangers) might do, rather than on managerial ability per se.

Regarding management practices, Bloom and Van Reenen (2007) began an important research program. Their initial paper reports results from an establishment-level survey of management practices in 732 manufacturing firms located in four countries. They present data on 18 management practices (including the degree to which the firm uses high-powered incentive systems, teamwork, selective recruiting, and skills training) and show that measures of management practice are highly correlated with TFP, profitability, Tobin's Q, sales growth and survival rate. They show that although 42% of the overall variation in management practices can be ascribed to country and/or three-digit industry effects, the remaining 56% is within country and industry. Furthermore, they demonstrate that management practices can explain a substantial portion of the variation in the permanent component (estimated from longitudinal data on the production function) of firm productivity.

Since their initial paper, Bloom, Van Reenen, and co-authors have substantially deepened this program, including expanding the data set to include more firms, more countries, and more kinds of organizations (Bloom et al, 2012b); expanding the conception of management to include organizational structure (Bloom et al., 2012d), and exploring the connection between management and information technology productivity (Bloom et al., 2012c). In addition, they

have made progress on identification issues, both through an instrumental variables approach (Bloom et al., 2010) and a field experiment (Bloom, et al., 2012a). Finally, they have summarized their work in various outlets (Bloom and Van Reenen, 2010, 2011).

Given all this work, one might wonder whether anything else has been or could be done. Our answer is strongly in the affirmative to both questions, for reasons like those described in the Ichniowski and Shaw (this volume) discussion of insider econometrics. Specifically, although large samples can offer important strengths, focused samples can offer complementary strengths, such as: (1) more detailed controls, (2) sharper performance measures, and (3) more nuanced understandings of what managers actually do. In the econometric framework $Y = X\beta + Z\gamma + \varepsilon$, these strengths correspond to Z, Y, and X, respectively. For our purposes, the first two are important, and the third is crucial: we perhaps could not have conceived of this chapter without the detailed accounts of management practices given by the kinds of focused-sample research we describe below. This is one of the chief returns that focused-sample researchers sometimes get from their days in the field talking with dozens of sources.

As a complement to large-sample research, this section describes focused-sample work.[3] In Section 3.1 we focus on blue-collar work in general and on "high-performance work systems" in particular; in Section 3.2 we turn to studies that have explored the role of management practices in driving productivity differentials in white-collar work. Throughout this discussion, we are illustrative rather than exhaustive, and we pay special attention to studies that illustrate the benefits of focused samples in providing detailed control variables (Z), sharper performance measures (Y), and nuanced measures of management practices (X).

Any interpretation of these results in both the blue-collar and white-collar processes literature must be tempered by the recognition that the adoption of any particular configuration of organizational practices is almost certainly not a random event. On the contrary, the endogeneity of (and documented complementarity among) these organizational practices naturally limits any claims of causality. Nevertheless, we take two important findings from these

[3] Indeed, the field experiment by Bloom, et al. (2012a) has many attributes of a focused study and so could be included here, but see Ichniowski and Shaw [this volume] for a description.

carefully conducted studies: (1) there is substantial variation in internal practices and (2) there is significant correlation between internal practices and performance.

3.1 Blue-Collar Work: High-Performance Work Systems and Productivity

A long tradition among labor economists and others studying blue-collar work has suggested that "high-performance work systems" are closely correlated with superior organizational performance (e.g., Kochan et al. 1986, Macduffie 1995, Huselid 1995, Huselid & Becker 1996, Ichiniowski et al. 1997, Pfeffer 1998, Appelbaum 2000, Black abd Lynch 2001).[4] For example, a large literature has developed around the study of Toyota, a firm that was perceived for many years to be an order of magnitude more effective than its competitors.[5] Similarly, Southwest Airlines has been widely studied as an example of best practice (see Gittell 2003). Both firms have significantly outperformed their competitors, and in both cases the difference has been widely ascribed to the ways in which they manage their blue-collar labor force.

There is no single definition of "high-performance work system," but three overarching elements can be identified in the literature. In general, firms with high-performance work systems (1) implement effective incentive systems, (2) pay a great deal of attention to skills development ,and (3) use teams and create widespread opportunities for distributed communication and problem solving. Table 1 lists some of the particular practices identified by some key studies in the field.[6]

---

[4] Regarding this distinction between large- and focused-sample studies, it is worth noting that Huselid (1995), Huselid and Becker (1996), and Black and Lynch (2001) are in fact large-sample studies of work practices, similar in some ways to the Bloom–-Van Reenen research on management practices. Huselid and Becker (1996) find that a one standard deviation increase in the use of high-performance work systems is associated with a $15,000 increase in market value/employee. Black and Lynch (2001) find that unionized establishments that adopted human -resource practices promoting joint decisionmaking coupled with incentive-based compensation had significantly higher productivity than did other similar nonunionized plants, whereas unionized establishments that maintained more traditional labor-management relations had lower productivity.

[5] More than 300 books and more than 3,000 articles have been published about Toyota and managerial systems. See for example, the international bestseller by Jeffrey Liker (2004), *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, and the companion book *The Toyota Way Fieldbook* (with David Meier, 2006) which is targeted toward managers trying to implement the 14 principles in their own companies.

[6] Becker and Gerhart (1996) provide a similar meta-summary of the practices included in additional studies of high-performance work systems.

Table 1: Management practices underlying high-performance work systems

Practice[7]	Appelbaum (2000)	Ichniowski et al. (1997)	Bloom and Van Reenen (2007)	Macduffie (1995)	Pfeffer (1998)	Huselid (1995)	Black and Lynch (2001)
High-powered incentives							
Incentive pay	X	X	X	X	X	X	X
Employment security	X	X	X		X		X
Merit-based promotions	X		X			X	
Reduced status distinctions	X			X	X		
Performance review			X			X	
Skills Development							
Skills training	X	X	X	X	X	X	X
Selective recruiting	X	X	X	X	X	X	X
Flexible job assignment	X	X		X			
Dense communication and local problem solving							
Teamwork	X	X	X	X	X	X	X
Communication	X	X					X
Information sharing	X				X	X	
TQM/Process Control			X	X			X

The first of these categories, *high-powered incentives*, includes both financial incentives (e.g., employee stock-ownership plans, performance-based pay, piece rates, and group-incentive plans) and longer term psychological or cultural incentives (e.g., long-term job security; plausible promotion opportunities; reduced wage and status differentials between workers and

[7] These practices are not necessarily mutually exclusive. For instance, some authors specifically focus on the importance of communication and information sharing in developing well-functioning teams, others directly measure the use of such teams, while still others measure a specific type of team that might be used in lean manufacturing, total quality management programs, or statistical process control. These may all be capturing similar successful work practices.

management; and pro-active conflict-resolution methods in place of autocratic, centralized control). The second, *skills development*, includes such actions as selective hiring processes, the extensive training of employees, job rotation and more flexible job assignments. Finally, *dense communication, local problem solving* includes such practices as the widespread use of teams; the decentralization of information gathering and processing; the allocation of decision rights to front line manufacturing employees (through self-directed teams, offline problem solving, or quality control groups); and improved communication and coordination among different employee work groups so that they have access to resources that may assist problem solving. Several studies have found widespread variance in the adoption of such practices (e.g. see Huselid 1995).

Studies that focus on specific industries allow researchers to develop industry-appropriate productivity measures and to collect more detailed and precise data, often at the plant or even line level. Focusing on a single industry also allows researchers to control for heterogeneity, while the smaller scale of these studies mitigates selection and response-bias problems. For example, Lazear (2000) studied workers installing auto windshields, finding that those who were under a piece-rate system were significantly more productive than those who were not. The change to a piece-rate system led to a 44% increase in worker productivity. About half of the increase in productivity appeared to be purely an effect of the greater incentives provided by the piece-rate mechanism; the remaining increase in productivity was due to a selection effect. See Ichniowski and Shaw (this volume) and Lazear and Oyer (this volume) for details.

In a global study of automotive assembly plants between 1987 and 1990, Macduffie (1995) contacted 90 assembly plants in 16 countries representing 60% of the industry's production capacity. He found statistically significant correlations among three sets of high-performance practices and productivity and quality: work system design, corporate level human resource management policies, and the (minimal) use of inventory buffers. Their *"work systems"* variables capture measures of communication and local problem solving, with additional focus on specific management innovations such as Total Quality Management and statistical process control; their "human resource management policies" variable overlaps heavily with work practices associated with improving workers' skills and incentives; and their "inventory buffers" variable is specific to the lean manufacturing methods associated with the Toyota Production

System. The nature of the capital stock, scale, model mix, parts complexity, and product design age together explained 48.3% of the variation in labor productivity, but adding their three "management practice" variables increased the adjusted R^2 to 54.3%, and adding in two- and three-way interaction effects further increased this measure to 64.9%. Subsequent work (MacDuffie et al.1996) showed that these practices diffused only slowly: in the 43 plants common to both the 1989 and 1993 surveys the overall 100-point work practices index increased from only 34.6 to 46.9.

Another example of an intra-industry study of the value of high-performance work systems comes from the work of Ichniowski, Shaw, and their co-authors in the steel industry (Ichniowski et al. 1997, Ichniowski and Shaw 1999, Gant et al. 2002, Boning et al. 2007). Ichniowski et al., (1997) examined the effect of human resource management practices on the productivity of integrated steel finishing lines. The authors collected data from 36 homogenous steel-finishing lines to form a panel data set of 2,190 monthly observations. Discussion with engineering experts and careful controls for the technology and product mix of any particular line led the authors to focus on "uptime" as a particularly clean measure of productivity. The study focuses on innovative practices in seven areas of personnel management: incentive pay, recruiting and selection, teamwork, employment security, flexible job assignment, skills training and communication, each measured using a range of qualitative variables. For example, the presence of teamwork and communication were identified by asking the questions listed in Table 2:

Table 2: Sample questions used to identify teamwork and communication

Variable name	**Dummy variable question**
Teamwork	
High participation	Are a majority of operators involved in formal or informal work teams or other related problem-solving activities?
Multiple teams	Do operators participate in more than one problem-solving team?
Formal team practice	Are operators organized into formal work teams either on the line or for the purposes of problem-solving activities according to an established policy with at least some operators involved in team activities?
Communication	
Information sharing	Are operators and union representatives, if any, provided with financial information on a regular basis?
Meet workers	Do line managers meet off-line with operators to discuss issues of concern, including issues related performance and quality?
Meet union	Do union representatives and managers meet often to discuss concerns and cooperate in finding solutions to issues?

Source: Table 1 in Ichniowski et al. 1997

The authors define four distinct systems of work practices. System 4 is the traditional system and lacks most of the innovative work practices. System 3 includes some low to moderate levels of teamwork and improved labor-management communication. System 2 includes the innovations of system 3 with higher levels of teamwork and also introduces extensive skills training. Finally system 1 incorporates all eight innovative work practices described by the authors. The authors find strong evidence that human resource management practices exist in bundles, and they use cross-section and fixed-effects specifications to suggest that it is bundles – *and not individual practices* – that have an economically and statistically significant effect on productivity.

Their data suggest that relative to system 4, systems 3, 2, and 1 improve line productivity by 1.4%, 3.2%, and 6.7% respectively.[8] Using detailed cost data from one line, the authors estimate that the 6.7% productivity difference between systems 1 and 4 implies increased profits of $2.24 million annually. In an attempt to identify causality, the authors also use a fixed effects specification to explore productivity changes within a particular line following the adoption of new work practices. Their data show that, lines upgrading from system 4 to systems 3 and 2 improved their productivity by 2.5% and 3.5%, respectively.[9] Controlling for selection effects (because lower performing lines are more likely to switch regimes) raises these numbers to 4.3% and 6.8%. Here again the data suggest that "better" practices are surprisingly slow to diffuse: 36% of the monthly observations in the sample used the traditional system for all 5 years and 58% still had no teams by the end of the sample.

Hamilton et al., (2003) used weekly productivity and personnel data for employees at a garment factory between 1995 and 1997 to investigate the effects of the introduction of a team-based production system on productivity. Between 1995 and 1996, both team and individual production methods were employed side by side under the same management and union, sewing essentially identical garments. They found that the adoption of teams over a 3-year period at a

[8] These estimates come from a regression using detailed technology controls, including customer type, maximum line speed, maximum width of line, nine dummies corresponding to specific equipment used in the line, equipment age, computer control of operations, and a measure of the value of major new equipment.

[9] All lines classified as system 1 are new lines and hence could not be included in analyses of line changes over time.

single plant was associated with an average increase in productivity of 14% (4 % stemming from higher ability workers more likely to join teams and 10% attributable to team production effect).

3.2 Management Practices in White-Collar Work

An extensive stream of empirical work has explored the implications of variation in white-collar work processes on productivity. However, in contrast to research on the blue-collar workforce, much of the work in this tradition abstracts from questions of incentives and skills to focus primarily on questions of decision-making and integrated communication.

For example, Clark and Fujimoto (1991) and Iansiti and Clark (1994) draw on detailed data collected at the level of individual product development projects to show that product development productivity in the world automobile industry between 1981 and 1988 was strongly correlated with the use of "heavyweight project teams" – an organizational structure in which the senior team leader has formal authority over the members of his or her team and is responsible for supporting rich communication and rich problem solving processes across functional groups. The authors show that the use of heavyweight project teams is correlated with higher development productivity, faster lead times, and higher quality projects. Successful firms also do simultaneous product and process engineering, and they rely on bilateral and face-to-face communication, and the constant and early release of information. For example, these show that Japanese firms (which were in general more likely to employ these techniques) took an average of 24 months to develop a new vehicle, whereas American firms (which were the least likely to use them) took 34.

In a study of R&D productivity in the pharmaceutical industry, Henderson and Cockburn (1994) used qualitative research methods to develop measures of management practice and study their effects on research productivity. Using data from structured interviews and internal company documents, the authors created the variable PROPUB to measure the use of incentives designed to persuade researchers to "act almost as if they were academic scientists" and to publish in the research literature. They created a second management variable, DICTATOR, to measure the degree to which decision making about resource allocation was concentrated in a single individual. The authors show that adding PROPUB and DICTATOR to their regression

predicting research productivity increased the R^2 statistic by 34% (from 0.49 to 0.66).[10] Despite the apparent power of these practices they diffused only slowly across their sample (Cockburn et al, 2000).

Researchers in health-care delivery have also increasingly examined organizational factors as a potential explanation for the wide variability in the adoption of clinical best practices and organizational performance. For example, in 2000, a highly motivated anesthesiologist from the Johns Hopkins University School of Medicine named Peter Pronovost set out to reduce medical errors in intensive care units (see Pronovost and Vohr 2010). He chose to focus initially on central line catheter-related bloodstream infections. Existing research had shown that consistent compliance with a few basic safety procedures could significantly reduce the rates of this type of infection. Another contributing factor in the study choice of catheter-related bloodstream infections was the large personal cost to the patient (morbidity and risk of mortality) and medical costs associated with these events (estimated annually in the United States to be 80,000 infections, 28,000 deaths, and $2.3 billion in additional costs). Finally, informal observation of clinicians in surgical intensive care units at Johns Hopkins suggested that these basic safety procedures were not in use.

Over time, based on his experience at Johns Hopkins, Pronovost developed a set of complementary management practices designed to reduce errors (specifically, catheter-related bloodstream infections) in hospital intensive care units. This set of management practices included (1) a checklist of five steps to be taken when inserting central line catheters, (2) organizational tools to support continuous quality improvement (e.g., performance data collection and feedback), and (3) a culture change program that was delivered in house and was designed to foster collaborative teamwork between clinicians and non-clinicians in reducing errors. In 2003, Pronovost and colleagues entered into a research collaboration involving nearly all acute care hospitals in Michigan (108 intensive care units). They designed and executed a pre- and post-experiment involving the three-pronged intervention. Over 18 months, median (mean) catheter-related blood stream infections were reduced from 2.7 (7.7) to 0 (1.4) infections per 1000 catheter-days (Pronovost et al. 2006). These reductions were sustained at 36 months

[10] This analysis included a wide variety of program level controls for productivity including resources spent by program, therapeutic area, and the scale and scope of the research effort as a whole.

post intervention (Pronovost et al. 2010) and were associated with significant decreases in hospital mortality (Lipitz-Snyderman et al. 2011).

4. Many Important Management Practices Are Relational

The evidence that heterogeneity in management practices may be one important source of persistent productivity dispersion among seemingly similar enterprises raises an important question: why, then, do best practices not diffuse? After all, changing management practices would seem much less costly than, say, upgrading capital equipment or buying the latest information technology.

As outlined in Section 1, we believe that one possible answer to this question may be that, first, the implementation of many competitively significant management practices requires a relational contract (again, roughly, a shared understanding of each party's role in and rewards from achieving cooperation) and, second, that such contracts are hard to build and change. Here in Section 4 we use field research and case studies to buttress the first part of this suggestion: we describe particularly successful blue- and white-collar work practices that rely on relational contracts. In Section 5 we then argue that these relational contracts can be hard to build and change, again drawing on qualitative evidence, this time partly from firms that had significant difficulty implementing competitively significant management practices.

In some sense, we have made our tasks in Sections 4 and 5 harder, because we allowed our descriptions of management practices in Section 3 to be deceptively precise, such as "commit to employment security," "develop teamwide communication," and so on. Such blithe descriptions can give the impression that management practices are like light switches—choices that can be easily switched on or off—so that a manager's greatest challenge lies in designing practices, rather than in implementing them. Here we argue that this view is mistaken and that many (most?) competitively significant management practices require both managers and employees to act in ways that cannot be fully specified ex ante or verified ex post, so organizations must rely on relational contracts to implement these practices.

One concrete example of this idea is provided by the case of Nordstrom (Spector and McCarthy 2012). For many years Nordstrom's employee handbook was a single sheet of paper on which was written:

Welcome to Nordstrom

We're glad to have you with our Company. Our number one goal is to provide outstanding customer service. Set both your personal and professional goals high. We have great confidence in your ability to achieve them.

Nordstrom Rules: Rule #1: Use good judgment in all situations. There will be no additional rules.

Please feel free to ask your department manager, store manager, or division general manager any question at any time.

Throughout this period Nordstrom was an extraordinarily successful firm, plausibly because their employees were extremely good at "exercising good judgment in all situations." Nordstrom's sales associates have reputedly accepted the return of snow tires (Nordstrom does not sell snow tires), driven for hours to deliver merchandise so that customers can be ready to attend family occasions, and changed the tires of customers stranded in the company parking lot. These actions, and others like them, have given Nordstrom a reputation for excellent customer service that is the envy of its competitors and that has created deep customer loyalty. Notice, however, that "exercising good judgment" is not something that can be easily defined. By its very nature it is not an action rule that can be readily described ex ante or verified ex post. Instead it is a process of noticing what this particular customer needs at this particular time and then deciding to take a particular action in response. No formal contract could fully specify what it means to "exercise good judgment" and it therefore seems plausible that the firm relies on relational contracts instead.

In this section we attempt to extend this insight and build the case that many—perhaps most—competitively significant management practices rely on relational contracts. In Section 4.1, we lay some formal groundwork by describing the repeated trust game from Kreps (1990) and briefly discussing relational-contract models that have built on this intuition. We then

present three qualitative examples of management practices supported by relational contracts. In Section 4.2, we discuss two blue-collar examples—the bonus system at Lincoln Electric and the use of the andon cord in the Toyota production system. And in Section 4.3, we discuss a white-collar practice—the nature of "pro-publication" management in the pharmaceutical industry.[11]

4.1 Modeling Relational Contracts

Repeated-game models of collusion among firms have been explored since at least Friedman (1971). The repeated trust game from Kreps (1990) has the same underlying intuition but is cast in a setting closer to our own and so allows us to develop some basic terminology. In Kreps's analysis, each period begins with player 1 choosing either to Trust or Not Trust player 2. If player 1 chooses Trust, then player 2 can choose either Honor or Betray; if player 1 chooses Not Trust, then the period ends. The payoffs to players 1 and 2 are (C, C) following (Trust, Honor), (S, D) following (Trust, Betray), and (P, P) following Not Trust, where $D > C > P$. That is, in the language of the prisoners' dilemma, D is player 2's payoff from defection (betraying 1's trust), C is her payoff from cooperation (honoring 1's trust), and P is her payoff from punishment (when 1 chooses not to trust). For simplicity, we assume that player 1's payoffs are the same values C from cooperation and P from punishment, but all that really matters is that 1's payoffs satisfy $C > P > S$ (where, again drawing on the prisoner's dilemma, S is 1's sucker payoff, from having his trust betrayed). Finally, to focus on the case where cooperation maximizes the players' total payoffs, we assume that $2C > S + D$.

In a one-shot trust game, player 2 will choose Betray, because $D > C$, so player 1 will choose Not Trust, because $P > S$. In the repeated game, however, the parties may be able to sustain cooperation, provided player 2 is not too impatient. For example, if player 1 begins by playing Trust but switches forever to Not Trust if player 2 ever plays Betray, then it is optimal for player 2 to play Honor if the present value of the payoff sequence (C, C, C, …) exceeds the present value of the payoff sequence (D, P, P, …). If player 2's discount rate is r, then it is optimal for her to cooperate if $r < (C - P) / (D - C)$; in this case, one could say that the players have a shared understanding that is "self-enforcing." See Gibbons and Henderson (2012) for

[11] These descriptions draw on Gibbons and Henderson (2012), "Relational Contracts and Organizational Capabilites," *Organization Science*, by permission from the Institute for Operations Research and the Management Sciences, 7240 Parkway Drive, Suite 300, Hanover, MD 21076, USA.

details, modest enrichments, and evidence from both experiments and contracts between firms that such relational considerations matter in practice.

The repeated trust game allows us to understand the basic intuition behind relational contracts, but it does not allow the parties to make payments to each other. Because employment relationships are perhaps our chief setting of interest (and because, as we will see, allowing for such payments has an important impact on the analysis), henceforth we focus mainly on models where payments are feasible. (Strictly speaking, the possibility of such payments is what distinguishes relational-contract models from repeated-game models.) The first such models were by Bull (1987) and MacLeod and Malcomson (1989), who analyzed intuitive equilibria but conducted these analyses before the literature learned from Abreu (1988) to analyze *optimal* equilibria. That is, it is one thing to determine that a particular set of strategies (say, reverting to the static Nash equilibrium after defection) can sustain cooperation if the parties are sufficiently patient, but it is quite another to determine what is the best equilibrium outcome the parties can achieve if they are not especially patient. The latter equilibrium need not have the simple form assumed in the former equilibrium. Indeed, for repeated-game models, the optimal equilibrium can involve quite sophisticated punishment strategies; see Abreu (1988) and Abreu et al (1990).

Levin (2003) rebooted the relational-contracts literature by proving the striking result that the possibility of payments between the parties makes optimal equilibria quite simple in many relational-contracting models, as opposed to optimal equilibria in repeated-game models. For example, in the canonical repeated agency model, where both parties are risk-neutral and have deep pockets, suppose the principal promises to pay the agent a bonus b as a function of the agent's output y. Levin showed that the optimal equilibrium is stationary, in the sense that the bonus in period t is a stationary function of only the output in period t, $b(y_t)$. Furthermore, this bonus function $b(y_t)$ might itself be simple (as well as stationary across periods), such as paying no bonus if output is below a critical value y^* and a bonus of b^* if output is above y^*. See Malcomson (this volume) for much more on this subject.

The simplicity of optimal equilibria in the canonical relational agency model has facilitated the application of such models in many settings, including some of those described in Section 5.2 below. Furthermore, the rebooting of this literature has led to richer models in which the optimal equilibrium is not stationary, including some of those described in Section 5.3 below. Before

turning to such models, however, we first illustrate the relational contracts underlying some famously successful management practices.

4.2 Relational Contracts in Blue-Collar Settings: Lincoln Electric and Toyota

Lincoln Electric is an Ohio-headquartered manufacturer of arc welders and has relied successfully on incentive pay for more than 75 years.[12] Most workers earn a piecework wage and receive an annual bonus that generally averages half the worker's annual pay. The board decides on the company-wide bonus pool, and individuals then receive shares according to a merit rating based on four factors—dependability, quality, output, and ideas and/or cooperation. Because of its incentive plan, and perhaps also for other reasons, Lincoln has succeeded fabulously in its industry, including driving established firms, such as General Electric, out of the market.

At first glance it might appear that Lincoln's piece rate and bonus could easily be duplicated, but our simple description again masks the presence of several nuanced relational contracts between the company and its employees that are central to the effectiveness of Lincoln's incentive plan. For instance, the company encourages workers to continuously improve piecework processes, but it reserves the right to change piece rates when there is "a change in methods." A worker who develops a faster method for a particular piecework process thus has a clear incentive to keep this knowledge concealed. If the worker's output suddenly skyrockets, managers might notice and readjust piece rates, but the worker could also produce the same amount of output in less time and with less effort while maintaining the outward appearance of constant hours and output. Both workers and managers might be better off in the long run if workers were to share possible process improvements, but workers are likely to share their knowledge only if they believe that they will receive some "fair return" from doing so. Thus arises Lincoln's relational contract about when and how to change the piece rate.

The annual bonus also has important discretionary components. Although it may be possible to develop relatively concrete measures of some aspects of any particular worker's performance, the value of each individual worker's ideas and cooperation may be quite difficult to quantify. To the degree that Lincoln wishes to encourage exactly this kind of less quantifiable behavior, any effective incentive contract will include a significant discretionary component.

[12] This account draws on Fast and Berg (1975).

Furthermore, even if the merit rating on which the annual bonus is based were to be constructed from purely objective measures, the board ultimately sets the firm-wide bonus pool. Although the board makes use of a bonus reserve to minimize variance in actual bonuses paid, the average bonus has still ranged from 78% to 129% of wages since 1945. In short, the board ultimately has complete discretion in setting the bonus pool, so workers must rely on a relational contract that the board will, indeed, pay the expected bonus. Table 3 summarizes some elements of the relational contracts that we believe to be in place at Lincoln.

Table 3: Relational Contracts at Lincoln Electric

	Action		
Agent	Cooperate	Defect	Punish
Worker	1. Work on improving piecework process 2. Share ideas and improvements	1. Shirk on process improvement effort 2. Keep improvements hidden from co-workers and managers	1. Stop or reduce process innovation effort 2. Keep ideas and innovations secret inhibiting further innovation 3. Quit
Manager	1. Adjust piece rate only when there is a genuine change in methods 2. Credit workers for ideas in assessing merit ratings 3. Set total bonus pool for all workers large enough to reward workers' contributions	1. Claim there is a change in methods to adjust for workers who have innovated new methods 2. Deny workers good merit ratings. 3. Limit the cumulative total bonus pool despite obvious contributions	1. Lower workers' merit ratings (and thus lower bonuses) 2. Fire workers

A similar reliance on relational contracts is evident in our next example, from Toyota's Production System (TPS). For many years Toyota's performance consistently exceeded that of the rest of the automobile industry.[13] As we outlined in Section 3, much of this performance differential appears to have been driven by the firm's management practices, rather than by any structural advantage. Indeed, Toyota's success came in the face of competition that possessed

[13] This account draws on Liker and Franz (2011).

many significant assets—including well-established brands and distribution channels, extensive technical knowledge, and in some cases local government support.

Extensive research has demonstrated that one important source of the extraordinary levels of productivity and quality obtained by Toyota in the late 1980s and the 1990s was the firm's ability to make continuous improvements to its production processes. Shop-floor workers were grouped into problem-solving teams that were encouraged to identify bottlenecks and inefficiencies in the production process and to explore potential solutions to them. Workers were asked to be alert for potential opportunities for improvement and to be creative and innovative in their search for solutions. Successful participation in these teams relied on a set of behaviors that could not be articulated ex ante or verified ex post, and hence could not be rewarded through a formal contract.

The use of the andon cord as a tool to improve the production process is a particularly compelling example. Every worker on the Toyota production line has access to an andon cord, a rope that can be pulled to alert a supervisor to the fact that the worker believes there is a problem on the line. Once the andon cord is pulled, if the supervisor cannot resolve the problem that the worker has identified in a reasonably short time, the entire production line may be brought to a halt. This event is enormously costly. Giving individual workers the power to stop the line on the basis of something as nebulous as "whenever you see a problem" implies, we believe, the existence of a relational contract between workers and the firm. Any worker pulling the andon cord must exercise considerable judgment in identifying potential problems, and must believe that he or she will not be penalized for potentially stopping the line. No formal contract can specify the conditions under which stopping the line is an appropriate thing to do. Table 4 sketches some key elements that may underlie the relational contract that supports the use of the andon cord at Toyota.

Table 4: Cooperation, defection, and punishment in the use of the andon cord at Toyota.

	Action		
Agent	Cooperate	Defect	Punish
Worker	1. Pull the andon cord when worker sees a problem 2. Offer suggestions on improvements to the production process (that might make workers' job redundant)	1a. Never pull the andon cord (out of fear of being punished) 1b. Pull the andon cord to stop the line and avoid work when there is no true problem 2. Keep improvements hidden from co-workers and managers	1. Sabotage the manufacturing line 2. Pull andon cord frequently 3. Engage in absenteeism
Supervisor	1. Recognize potential problem when andon cord pulled and aid in problem-solving 2. Implement improvements without necessarily cutting jobs 3. Accept authority of work teams to make some shop-floor decisions	1. Punish workers for pulling andon cord (even appropriately) 2. Cut workforce once they discover potential innovations 3. Interfere in work teams and override their decisions	1. Penalize workers (financially or socially) for pulling andon cord 2. Remove the andon cord

4.3 Relational Contracts in White Collar Work: Pro-Publication at Merck

A similarly complex relational contract has been operational for many years, we believe, in some leading pharmaceutical firms.[14] For many decades, research-based pharmaceutical firms attempted to discover new drugs through a large-scale process of trial and error, so called random drug discovery. For example, researchers might inject hundreds of compounds into rats or dogs in the hope of finding one that lowered cholesterol or reduced hypertension. If a compound showed promising activity, then medicinal chemists would synthesize large numbers of related compounds for further testing, in an attempt to find potential drug candidates that were both more efficacious and less toxic. Highly skilled medicinal chemists were the backbone of most companies' research efforts.

[14] This account draws heavily on Henderson (1994)

Beginning in the 1980s, university-based researchers working in biology and biochemistry began to uncover the biochemical mechanisms underlying many diseases. Early work exploring the causes of hypertension, for example, uncovered the fact that blocking the renin angiotensive cascade lowered blood pressure. This discovery, and others like it, enabled firms to move from animals to using chemical reactions as screens to discover new drugs. In place of "find me something that will lower blood pressure in rats," firms could request "find me something that inhibits the action of the angiotensin II-converting enzyme."

This science based approach promised to be—and, ex post, proved to be—much more productive than the approach it replaced, but it required those firms who embraced it to significantly change the ways in which they managed research. First, the firms had to hire scientists with quite different sets of skills, including molecular biologists, physiologists, and biochemists. Second, whereas random drug discovery required the firm's scientists to maintain only a cursory understanding of the current scientific literature, the science-based approach required them not only to stay current with the promising mechanisms being discovered in the academic literature but also to conduct such leading-edge science themselves, in house.

This shift required research scientists to behave almost as if they were academic scientists. On the one hand, staying current with the academic literature or, more importantly, *understanding* the current literature, could not be accomplished by simply reading journal articles. In a classic instance of what Cohen and Levinthal (1990) have called "absorptive capacity," researchers wishing to stay at the leading edge found that they not only needed to attend academic conferences but also to conduct leading-edge science themselves (e.g., to be invited to such conferences, or to particpate in informal discussions at them). Some pharmaceutical firms even began to encourage their scientists to publish in the peer-reviewed literature—even though protecting the firm's intellectual property through patents was a substantial part of the ultimate goal.

At the same time, scientists in science-based pharmaceutical firms could not act completely like academics. They also had to actively support their colleagues in the attempted discovery of new drugs. Making a major scientific breakthrough—even winning a Nobel prize—without simultaneously generating knowledge that could be quickly applied to the search for new drugs would not constitute consummate success. Instead, those firms that chose to adopt the

science-based approach—and the scientists who worked for them—had to learn how to balance the behaviors characteristic of a successful research scientist with those characteristic of a successful drug hunter.

This was not a balance that could be easily specified in a formal contract. It was widely understood that staying connected to the academic world required publishing in journals and attending conferences, but no one could specify ex ante exactly how many papers a scientist should publish and which conferences he or she should attend, particularly when every scientist's research trajectory was likely to be different. How could anyone tell, ex ante, exactly what the balance between fundamental research and drug hunting should be at any given moment, or develop a verifiable measure of whether a researcher was being "sufficiently helpful" to his or her drug-hunting colleagues? This uncertainty was symmetric: scientific researchers might be motivated as much by the chance to work on interesting problems and to control their own time as by money or promotion, but specifying in advance exactly what kinds of rewards would be given in response to exactly what kinds of behaviors proved to be very difficult.

In practice, those pharmaceutical firms that successfully developed pro-publication management (as Cockburn et al. (2000) have termed it) not only fared better in the production of major patents (Henderson and Cockburn, 1994) but also appear to have developed a sophisticated relational contract with their researchers. Table 5 sketches our sense of its key elements.

Table 5: Cooperation, defection, and punishment in science driven drug discovery

	Action		
Agent	Cooperation	Defection	Punishment
Scientist	Behave almost like an academic scientist, but be sure to develop useful knowledge for discovering new drugs.	Either shirk (represent lack of results as unlucky research) or behave like an academic scientist (pursue problems for their own sake, build external reputation)	Behave like an academic scientist, or ignore research and become a drug hunter
Manager	Reward the scientist who displays high-science behavior even if no new drugs result.	Fail to increase resources for scientists who publish; reward only those who produce drugs	Fire the scientist, or cut funding

Taken together we hope that these three examples make plausible the idea that many—perhaps most—competitively significant management practices rely heavily on relational contracts for their execution. Such shared understandings ask both managers and employees to perform actions that cannot be easily specified ex ante or verified ex post, and as such cannot be enforced by formal contracts.

5. Could Relational Contracts Be Part of the Implementation Problem?

It is one thing to assert that competitively significant management practices may rely on relational contracts, but quite another to suggest that this dependence is one of the reasons that such practices are slow to diffuse and hence that performance differences persist. To believe the latter, we would need to know why the relational contracts underlying key management practices might themselves be slow to diffuse. In this section and the next we discuss three approaches to exploring this possibility.

To put these three approaches in context, we begin in Section 5.1 by briefly summarizing existing explanations for the slow diffusion of management practices—without reference to relational contracts. As noted in the introduction, this existing literature has proposed four broad answers to the question of why competitively significant management practices do not diffuse more rapidly, including the following. First, incumbent managers may have problems of *perception*—they do not know they are behind. Second, managers may have problems of *inspiration*—they know they're behind, but they don't know what to do about it. Third, managers may have problems of *motivation*—they know they're behind and they know what to do, but they lack incentive to adopt new practices. Fourth, managers may have problems of *implementation*—they know they're behind, they know what to do, and they're trying hard to do it, but they nonetheless cannot get the organization to get it done.

The first of our three approaches to understanding why some relational contracts might be slow to diffuse is related to the motivation problem: if the relational contracts that facilitate high performance in some enterprises are simply too costly or not feasible for others to adopt, then low-performing enterprises will know they're behind and know what could be done, but they will find it unprofitable or infeasible to do it. In this scenario, high- and low-performing enterprises may seem similar but actually differ in some important parameter. In short, this

approach is consistent with the idea from Section 4 that high-performance management is often importantly relational, but this approach then attributes PPDs to unmeasured heterogeneity in the costs of adopting the key relational contracts.

We discuss models of this kind in Section 5.2. In particular, we describe models with four sources of heterogeneity—differences in discounting, in the competitive environment, in network structure, and in player types—that might lead to heterogeneity in the kinds of relational contracts that firms can adopt. These models offer interesting applications of the basic logic from Levin's (2003) canonical model, and they suggest further topics that could be explored with related techniques.

On the other hand, the models in Section 5.2 cannot account for the important class of cases in which: managers believe that the adoption of a particular management practice would improve productivity, they have a deep understanding of the nature of the practice, they are striving mightily to adopt it, and yet they still fail to implement the practice successfully. In the introduction we mentioned the mature phases of the Toyota Production System and Total Quality Management as examples from this class; now, after Section 4.2, we can add science-driven drug discovery such as at Merck to the list.[15]

Our second and third approaches to understanding why some relational contracts might be slow to diffuse both explore implementation difficulties that arise through relational contracts. In particular, our second approach includes a variety of recent models that enrich Levin's canonical model and deliver non-stationary optimal equilibria. In contrast, our third approach is more speculative and has not yet been modeled, as far as we know.

We discuss the second approach in Section 5.3 by describing models such as those where path dependence in relational contracting can produce persistent performance differences among ex ante identical enterprises. As noted in the introduction, several interesting possibilities arise: cooperation can be hard to build well, in the sense that achieving perfunctory cooperation can

[15] For example, there have been more than 300 books and thousands of articles written about Toyota, yet until quite recently many automobile companies appeared to have great difficulty imitating its practices (Pil and MacDuffie; 1996). Similarly, the practice of science-driven drug discovery diffused only slowly across the pharmaceutical industry despite widespread agreement about its effectiveness (Cockburn, Henderson & Stern, 2000).

make it harder to achieve consummate cooperation; cooperation, once built, can be fragile; and cooperation may be difficult to build in the first place. In all these models, however, the parties play the optimal equilibrium of the game as a whole, and they understand this equilibrium from the beginning of the game. Thus, although we have moved beyond the bad parameters that cause bad performance in Section 5.2, now we have explanations of bad performance based on bad luck—i.e., from path dependence or other events over the course of the relationship, among enterprises that were identical ex ante.

Finally, because our third approach is more speculative, with no relational-contract models we know of, we discuss it separately, in Section 6. Indeed, we begin not with theory but instead with brief descriptions of cases where firms had difficulty building or changing relational contracts. Based in part on these cases, we suggest that the implementation of some relational contracts requires solving problems of not only *credibility* but also *clarity*. The credibility problem is familiar: should one party believe another's promise? The clarity problem is new: can one party understand another's promise? In fact, the clarity problem is tougher than this statement, for the usual reasons associated with common knowledge: I need to understand that you understand my promise, and so on.

We emphasize that the clarity problem is likely to be particularly acute in exactly those settings where relational contracts are needed because formal contracts are imperfect or infeasible: where parties recognize that their roles in and rewards from cooperation are not easy to articulate ex ante or verify ex post. More precisely, the existing relational-contract models described in Section 5 (both those focused on bad parameters in 5.2 and those focused on bad luck in 5.3) share the focus of the entire relational-contracts literature on what happens when the parties' roles in and rewards from cooperation cannot be verified ex post. This literature's focus on what happens without verifiability ex post has been very productive; we think it is now time to start studying what happens when these roles and rewards cannot be fully articulated ex ante. We say more about how one might do this in Section 6.

5.1 Existing Explanations for the Slow Diffusion of Management Practices

An extensive literature has attempted to account for the fact that many competitively significant management practices do not diffuse rapidly. As noted above, four major lines of

argument involve the perception, inspiration, motivation, and implementation problems. For ease of exposition, we proceed from perception to motivation and then to inspiration and implementation.

5.1.1 The Perception Problem

A first stream of work hypotheses that significant heterogeneity in management practices persists because managers are subject to problems of perception. It argues that managers may literally not be aware that heterogeneity in management practices exists or that these differences shape productivity. For example, Henderson and Clark (1990) suggest that in the photolithographic alignment equipment industry, incumbent firms failed to recognize the threat posed by the next generation of products because their employees focused their attention on the components of each technology, failing to understand that their competitors had developed the ability to understand the performance of the system as a whole. Christensen (1997) suggested that incumbent firms in the disk-drive industry invested their attention and resources into satisfying the needs of their existing customers, and in doing so failed to appreciate the ways in which their competitors had developed skills that allowed them to meet new customer needs using new technology. Trispas and Gavetti (2000) drew on a detailed history of the competition between Polaroid and Kodak to show that the two firms conceived of photography in very different terms and that this led them to make very different investments. More recent work has amplified and extended these results, suggesting that both the cognitive frames of individual managers and the ways in which information processing is structured within the firm may lead firms to have quite different perceptions of their environments and the capabilities of their competitors (Kaplan et al. 2003, Kaplan 2011).

5.1.2 The Motivation Problem

A second stream of explanation has focused on the degree to which heterogeneity in the adoption of management practices persist because managers have motivation issues, either because of differences in competitive intensity across industries or because of difference in agency costs within firms. This perspective has most recently found support in the work of Bloom and Van Reenen (2007); see also Reinganum (1989) and Bresnahan, Greenstein, and Henderson (2011) for alternative formulations. Bloom and Van Reenen report statistically

significant correlation between the adoption of their measures of management practices and competition, measured by three separate proxies: import penetration, the Lerner index, and self-reported survey results on the level of competition. Increasing competition from "few" to "many" (with the cutoff being five competitors) is associated with an increase in their management z-score of 0.140. Bloom, Sadun, and Van Reenen (2012) find even stronger results for estimates with industry-country fixed effects in panel data.

5.1.3 The Inspiration Problem

The third stream of explanation for persistent heterogeneity in management practice flows from the seminal work of Sidney Winter and his collaborators, and suggests that many competitively significant management practices fail to diffuse because while managers may be aware that their competitors have developed practices that it would be valuable to adopt, they are hampered in adopting new practices by an inspiration problem: much of the knowledge fundamental to building management practices is "tacit," or deeply embedded in individuals or organizational routines, so it is simply unclear how to mimic a high-performer (see, for example, Winter (1987, 1988, 2006); Nelson and Winter, 1982; Zollo and Winter, 2002).

Winter (2006) likens the problem of transferring organizational knowledge to the difficulty of baking a cake from a recipe: "'knowing how to bake a cake' is clearly not the same thing as 'knowing how to bring together in one place all of the ingredients for a cake' (p. 131)". As to the sequence of actions required to turn the ingredients into cake, he notes that while it is technically feasible to write out these tasks in great detail, hands-on experience under an experienced chef is a much easier and more effective way to learn to cook than any amount of book learning. This idea has been explored empirically by several researchers. Anand & Khanna (2000), for example, suggests that one of the reasons some banks are much better at making acquisitions than others is because there is a great deal of "causal ambiguity" in the knowledge required to make good acquisitions, and because some banks have greater "absorptive capacity" – i.e. more knowledge of the area in question, and so are better equipped to understand new ideas. Similarly a number of studies that have shown that both access to knowledge and performance are correlated with the movement of key individuals (See for example, Almeida and Kogut 1999; Breschi and Lissoni 2009; and Lacetera et al., 2004), suggesting that management

practices are much easier to "learn" if one has access to individuals who have experienced them first hand.

Detailed studies of particular management practices confirm the idea that many practices have important "sticky" or "causally ambiguous" elements. For example Gant, Ichniowski, and Shaw (2002) conduct a study of the drivers of innovation across four steel finishing plants. They find that more innovative plants are characterized by an emphasis on worker-level decision-making, team-based work, and greater communication in general with particular emphasis on horizontal communication between workers on the floor-shop. Their detailed diagrams of communication patterns show that innovative plants have more horizontal communication and denser interaction networks and suggest that once we examine "communication" at a sufficiently fine-grained level it is a highly complex set of behaviors that might indeed be hard to reproduce.

5.1.4 The Implementation Problem

Finally, introducing the implementation problem (and to some extent blending it with an inspiration problem), a fourth reason why it might be difficult to imitate a high performer is based on the idea that effective management practices are subject to strong complementarities; for example, see Milgrom and Roberts (1990, 1995) and Brynjolfsson and Milgrom (this volume). When the elements of a management practice are highly interactive, there can be little return from getting most of the elements right but large return from getting them all right. As a result, a firm could know that it is behind, know what it needs to do, be trying hard to do it, and yet not be able to get the many pieces of the organization sufficiently coordinated to get the job done.

As evidence in this spirit, recall from Section 3.1 that Ichniowski, Shaw and Prennushi's work suggests that high performance work practices are most tightly correlated with productivity when they are "bundled" together. Similarly, Pil and Macduffie (1996) find that firms are more likely to adopt what they deem "high involvement work practices" (use of teams, problem-solving groups, job rotation) when they previously had in place certain "complementary human resource practices" (selective hiring, contingent compensation, training, and reduced status barriers between management and production workers). Finally, Bresnahan et al. (2002) find

evidence of complementarities between the adoption of workplace organization (a measure of teamwork and decentralized work practices), information technology, and more skilled workers.

Levinthal, Rivkin and Siggelkow, amongst others, build on these observations to develop models of bounded search over rugged landscapes consistent with the idea that firms that search "locally" may find it very difficult to adopt new bundles of practices even under quite weak assumptions as to bounded rationality and costly search (Levinthal, 1997; Rivkin and Siggelkow, 2003; Siggelkow and Rivkin, 2005). Rivkin (2000), for example, suggests that it can be exceeding difficult to search through a space of strategies with numerous elements that are highly interactive, and that imitators may suffer large penalties from even small errors in attempting to match a particular combination of practices. Similarly Brynjolfson and Milgrom (this volume) suggest that the existence of complementarities plausibly makes change in management practice – when change must occur simultaneously across the entire bundle of practices if performance is to improve performance –costly or infeasible.

5.1.5 Summary

Taken together these four streams of research—on the perception, motivation, inspiration, and implementation problems—offer compelling explanations for why management practices may not readily diffuse. But none of these research streams has paid much (if any) attention to relational contracts. Given the suggestion from Section 4 that competitively significant management practices are often relational, in the rest of this chapter we begin to explore how relational-contract models might complement existing work in these research streams. In particular, in Section 5.2 we explore models related to the motivation problem: if the relational contracts that facilitate high performance in some enterprises are simply too costly or not feasible for others to adopt, then low-performing enterprises will know they're behind and know what could be done, but they will find it unprofitable or infeasible to do it.

In Sections 5.3 and 6, we then shift to our main focus—implementation difficulties that arise through relational contracts. We see our approach to the implementation problem as importantly complementary to the rugged-landscape models of Levinthal, Rivkin, and Siggelkow. More specifically, a firm that can engage in only local search may have difficulty on a rugged landscape, perhaps getting stuck at a local peak, whereas any firm that can hill-climb

locally can find the global peak of a landscape where all monotonically increasing paths lead to the top. But our canonical setting—say, something like Levin's relational agency model—is just such a smooth landscape, so local search would suffice. Instead, in our approaches, relational contracts pose implementation difficulties because the enterprise may have trouble moving at all (not to mention searching locally) because getting the organization to move requires the right relational contracts, and these may be hard to build.

5.2 Unobserved heterogeneity in the costs of using relational contracts

One relational explanation for persistent performance differences is that some enterprises find it too costly or infeasible to use the relevant relational contracts.[16] Here we draw from recent literature to explore four potential sources of such heterogeneity: differences in discounting, in the intensity of competition, in network structure, and in the "types" of managers or employees. This literature is making nice progress; our summary is designed not to be a comprehensive listing but rather to be suggestive of the potential fruitfulness of this approach.

Perhaps the most intuitive source of variation in the effectiveness of relational contracting is variation in discounting: impatient players are more tempted to defect. (Of course, a player's discount rate need not reflect her personal rate of time preference; it could instead reflect, say, the frequency of repeated interactions.) As one of many recent papers that produces this comparative-static result but is otherwise focused on different issues, consider Board (2012), where a firm builds a group of suppliers over time and then, once the group is established, sources each period from the low-cost provider from the group. A more patient firm optimally builds a larger group of suppliers and thus has lower expected procurement cost per period. An econometrician who could not control for the size of the supplier group or the firm's discount rate would therefore measure persistent differences in the firm's input costs that are created by the quality of its relational contracts with its suppliers.

[16] By focusing here and hereafter on models of PPDs sustained by relational contracts, we are certainly not rejecting alternative theories, including those related to the perception or inspiration problems. Indeed, we cannot resist noting some of the fascinating learning theories that could be seeds for the latter models, such as Smith and Sorensen (2000), Jeitschko and Taylor (2001), Bar-Isaac (2003), Bonatti and Hörner (2011), Callander (2011), and Ali and Kartik (2012).

As an illustration of how variation in competitive intensity could affect relational contracting, consider Powell (2012), who explores the interaction between relational contracts inside an organization and competition in the market outside. In particular, Powell builds on MacLeod and Malcomson's (1989) result that the key to sustaining a relational contract is the net surplus that the relationship creates over and above the parties' outside opportunities. (See Malcomson (this volume) for details.) Powell embeds many firms within one market, where each firm consists of a principal-agent relationship in which the principal creates incentives for the agent through a relational incentive contract. For a given firm, the net surplus from its principal-agent relationship depends on the market price of output, which in turn depends on the aggregate output produced by the other firms in the market. Loosely speaking, competition in the market may thus be bad for productivity in a firm, if a lower market price of output reduces the net surplus in the firm's agency relationship. More specifically, if market 1 consists of firms 1a and 1b and market 2 of firms 2a and 2b, then firms 1a and 2a might be internally identical and yet able to sustain different relational contracts—because of differences in the productivities of firms 1b and 2b.[17]

As illustrations of how variation in network structure could affect relational contracting, consider the recently emerging literature on repeated games in networks. This literature suggests that the network structure of interactions, the network structure of information, optimal network structures, and endogenous network structures may shape the kinds of relational contracts that can be constructed; see, for example, Lippert and Spagnolo (2011), Fainmesser (2012), Jackson et al. (2012), and Wolitzky (2012). This emerging literature may also allow important connections to empirical work such as the Gant et al. (2002) paper described in Section 5.1, which relates work-group networks to productivity. Our point here is simply that differences in network structure within firms might be hard to measure and hence a source of PPDs among enterprises that are only seemingly similar. A further step would be to see network structure as not only endogenous but perhaps also responsive to management practices.

[17] See also Board and Meyer-ter-Vehn (2011) for more on how industry equilibrium might affect relational contracting within individual firms. In particular, identical firms hiring identical workers in a given labor market may offer different relational contracts in equilibrium. Workers accept inferior contracts only until high performance gets them an offer from a firm offering a superior contract. Firms offering inferior contracts eventually lose their workers, but they profit in the short run from the high effort the workers supply in the attempt to generate an outside offer.

Finally, as an initial illustration of how private information about a player's type can affect relational contracting, consider Halac (2012), who adds persistent private information to Levin's (2003) analysis of relational contracting by endowing the principal with private information about her outside option (and hence about the value of the relationship, from which follows the maximum feasible strength of incentives for the agent). Unlike Levin's case of complete information, where it is immaterial whether the principal or the agent has the bargaining power, Halac derives different reneging temptations and hence different (and possibly testable) time-paths for incentive strength and termination probability depending on which party has the bargaining power. In particular, when the principal has the bargaining power, in the efficient equilibrium, incentives strengthen gradually, as the agent becomes convinced of the principal's type. Consistent with persistent performance differences, both of the principal's types in Halac's model value the relationship enough to stick with it, so after the bad type reveals herself, the relationship may not terminate but rather settle into perfunctory rather than consummate cooperation. See Watson (1999, 2002) for other models where relationships start small in order to screen types.

Another set of models endows the agent with persistent private information, rather than the principal. For example, in MacLeod and Malcomson (1988), workers know their own abilities, but initially firms do not. The technology of production is such that, under full information, it would be efficient for higher-ability workers to exert more effort. The result is akin to heats in a track meet. In the first round, only the least-able workers fail to qualify for the second round (*i.e.*, fail to get promoted); all other workers work just hard enough to qualify. Of those in the second round, the least able fail to qualify for the third and the others work just hard enough to qualify. Eventually, the continuous distribution of ability is broken into intervals, corresponding to the groups of workers who failed to qualify for each successive round. Having reached its steady-state level, each such group is playing an efficiency-wage continuation equilibrium—where the firm pays a high wage each period but will fire the worker if output is below a critical value. Of course, different groups of workers are playing (appropriately) different efficiency-wage continuation equilibria: a group with high average ability receives a higher wage and produces greater output than does a group with low, so an econometrician will measure persistent productivity differences if ability is not a control variable. See Malcomson (2012) and Yang (2012) for further models in this spirit. As with variation in network structure, a further step in

this literature would be to see the distribution of workers' types as responsive to management practices—whether via screening during hiring or skill development during employment.

5.3 Consummate Cooperation Can Be Hard to Build and Sustain

We turn next to our second approach to understanding why some relational contracts might be slow to diffuse. Whereas the relational-contract models in Section 5.2 bear some resemblance to the motivation problem—in that the relational contracts that facilitate high performance for some enterprises are simply too costly or not feasible for others to adopt, so underperformers know they're behind and know what needs to be done but are not trying to do it—the models we describe here are our first stab at our main interest: implementation difficulties arising from relational contracting.

In this section we focus on path dependence and other events during a relationship that can cause ex ante identical enterprises to perform differently. As will become clear, this literature is very recent but also very exciting. We hope it continues to grow.

In Section 5.3.1 we describe how learning about persistent private information potentially causes a principal-agent relationship to settle for perfunctory cooperation, when the principal knows that consummate cooperation is feasible but too expensive to discover. This model thus produces persistent performance differences among ex ante identical enterprises. In Section 5.3.2 we describe how learning about i.i.d. private shocks can cause non-stationary dynamics if the shock persists. In particular, in this model short-run shocks can produce long-run or even permanent distortions. Finally, in Section 5.3.3 we describe several models in which cooperation, once built, is nonetheless fragile. Of course, a publicly observed change in the payoffs can cause initial cooperation to collapse, but interesting recent models explore much more subtle threats to cooperation involving learning in various ways.

In all of these models, the parties play the optimal equilibrium and have a shared understanding of this equilibrium from the beginning of the game. In Section 6 we discuss our third approach to why some relational contracts might be slow to diffuse, this time exploring how clarity problems might hamper the parties' attempt to develop a shared understanding.

5.3.1 Persistent Performance Differences among Ex Ante Identical Enterprises

We now consider Chassang's (2010a) model of building a relationship. As in the types models in Section 5.2, such as Halac (2012) and MacLeod and Malcomson (1988), persistent private information inspires an initial learning phase. The key difference between the Chassang and types models, however, is that here the agent's private information is about production possibilities, not about an aspect of the agent himself, such as his ability. Our interest in this seemingly slight difference in modeling is that now the principal's attempt to learn and utilize the agent's information can produce persistent performance differences among ex ante identical enterprises (i.e., among enterprises with the same production possibilities), as follows.

Chassang shows that in the optimal equilibrium of the game as a whole, play converges to one of several heterogeneous steady states, based on the stochastic results of the initial learning phase. Furthermore, these steady-state continuation equilibria correspond to equilibria of the underlying game without private information. In this sense, Chassang's model can be interpreted as an equilibrium theory of equilibrium selection, potentially rationalizing Leibenstein's conjecture that under-performing enterprises might be stuck in Defect-Defect, whereas superior performers might have learned to play Cooperate-Cooperate.[18]

Formally, in each period, the Principal first chooses whether to invest or not, where investing imposes a cost k on the Principal but delivers a benefit b to the Agent. Not investing delivers zero to both parties and means that no further actions are feasible that period. If the Principal does invest then the actions $a \in A$ might be feasible for the Agent: each period, each action is feasible with (independent) probability p. Both parties observe which actions are feasible and the principal observes the action the agent chooses.

There are two kinds of actions, unproductive and productive: $A = A_U \cup A_P$. An unproductive action costs nothing for the Agent to take but produces no output for the Principal, whereas a productive action costs c to take and produces output $\tilde{y}(a)$, where $\tilde{y}(a) = y(a) > 0$ with probability q and $\tilde{y}(a) = 0$ with probability 1-q. It is common knowledge that the number of productive actions is $\#A_P$ and that a given productive action $a_p \in A_P$ produces $y(a_p)$ when it produces positive output, but initially only the Agent knows which actions are the productive ones.

[18] The remainder of this description of Chassang (2010) draws on Gibbons (2010) with permission from the *Annual Review of Economics*, Volume 2, © 2010 by Annual Reviews, http://www.annualreviews.org.

As a simple case, suppose it is common knowledge that there are two productive actions, a_0 and a_1, with $y(a_0) < y(a_1)$. In the first period, a_0 might be feasible but a_1 not, or the reverse, or both might be feasible, or neither. To induce the Agent to take a productive action that has not been revealed as such, instead of an unproductive action, the Principal threatens not to invest in several future periods if this period's output is zero. Note that this punishment will occur on the equilibrium path, because a productive action could produce zero output. In this sense, learning (i.e., identifying a new action as productive) is expensive. On the other hand, if an action has produced a positive output then the Principal knows that the action is productive, so if the Agent takes this action in a later period and it produces zero output then the Principal does not need to punish the Agent.

Because learning is expensive (and all the more so after at least one productive action has been identified), it can be optimal to stop learning before all productive actions are identified. Because opportunities to learn arrive randomly, otherwise identical dyads may stop learning after identifying different sets of productive actions. Thus, this model can produce persistent performance differences among ex ante identical dyads because of path-dependence in building a relationship.

5.3.2 Long-Run Distortions from Short-Run Shocks

In this section we consider Li and Matouschek's (2011) relational-agency model. Here the principal is subject to i.i.d. private shocks that increase her opportunity cost of making payments to the agent. The model then delivers three interesting results. First, the optimal equilibrium involves appealing economic logic concerning how bonuses, wages, effort (and hence productivity) evolve if the shock persists. Second, these predictions fare reasonably well in the limited case-study data available. Finally, when the shock makes the principal literally unable to pay, rather than just facing a higher opportunity cost of payments, the effect on the relationship can outlast the duration of the shock.

Formally, because the principal would always like to claim that it is difficult to pay, the optimal equilibrium imposes a cost on the principal for making such a claim. Li and Matouschek show that these costs come in the form adjustments to bonuses, wages, and effort as the number of consecutive periods of high opportunity cost increases. In particular, when the principal first experiences difficulty in paying, she keeps the agent's effort from falling too far by promising the agent a larger bonus come the day when payments are not so difficult. There are limits to this approach, however, and if the principal continues to have difficulty paying, she adds contractual wage payments to the relational bonuses. Because the principal faces a high opportunity cost,

however, these contractual wages also have their limits, so if the principal continues to confront a high opportunity cost, the contractual wages and the relational bonuses reach steady-state levels sufficient to induce some (but greatly reduced) effort from the agent, until the opportunity cost finally returns to its normal level.

Interestingly, Li and Matouschek find support for their predictions from two cases we will discuss in Section 6.1: Credit Suisse First Boston and (further history from) Lincoln Electric. For example, at CSFB, in line with this model, management's response in the first year of a shock was to promise superior bonuses when times got better, but their response in the second year was to offer guaranteed pay raises immediately. During this evolution of management's responses, the bankers did not switch abruptly from cooperation to defection, but they might well have decreased their effort, again in line with model.

Compared to some reputation models—where a persistent type might pick a particular moment to reveal herself, after which behavior may change discretely—here the principal's private information is i.i.d., but a run of bad luck can nonetheless lead to an interesting and realistic time-paths for bonuses, wages, and effort. Furthermore, Li and Matouschek show that when the principal is liquidity-constrained (ie, literally unable to pay, rather than just facing a higher opportunity cost of payments), the recovery of the relationship can be sluggish, with past bad shocks having a lingering or even permanent negative effect on the efficiency of the relationship.

5.3.3 Might Cooperation Be Hard to Sustain, Even Once Initiated?

As a stark example where cooperation collapses, imagine enriching the repeated Trust Game from Section 4.1 so that player 2's payoffs begin as (C, D, P) but each period there is a probability that 2's payoffs are publicly observed to have shifted permanently to (C, D′, P), where D′ > D and $r > (C - P) / (D' - C)$. In this case, cooperation will end after player 2's payoffs shift, but cooperation can nonetheless start at the beginning of the game, provided that r is sufficiently below $(C - P) / (D - C)$, relative to the probability that player 2's payoffs shift.

Richer versions of this idea explore how cooperation can collapse from transitory shifts in player 2's payoffs, or slower learning about permanent shifts in 2's payoffs, or other subtler causes. For example, McAdams (2011) models a stochastic partnership under symmetric information. There can be learning about the partnership's productivity, which can evolve, and the partners' actions can affect future productivity. A vicious cycle can arise where a negative shock bodes ill for the future, reducing the value of the ongoing relationship, possibly causing the partners to work less hard now, thereby further imperiling the relationship.

Whereas the state variable in McAdams' model is public information, Chassang (2010b) studies the effect of the global-games perturbation, where each party receives a low-noise, conditionally independent, private signal of the state. In particular, Chassang analyzes a dynamic exit game, where staying is akin to cooperation and exiting is akin to defection. The extreme equilibria—those with most staying and most exiting—involve threshold strategies, where a player stays for the next period only if her noisy signal is sufficiently high. In such equilibria, each player is uncertain about the other's behavior, and this makes it difficult to coordinate their expectations about play. Even as the noise goes to zero, the conditions needed for cooperation are stricter than they would be under complete information. Moreover, these conditions now depend on new considerations: now not only the predatory incentive (to defect when one's partner is cooperating) but also the pre-emptive incentive (to avoid cooperating when one's partner defects) matter for whether cooperation can be sustained.

As the first of two more applied examples, in Chassang and Padro i Miquel's (2010) related model of deterrence, weapons stocks are unambiguously helpful in deterring war under complete information, but they may be harmful with vanishingly little private information if they greatly reduce the payoff from being the second-mover into war, even if they also somewhat reduce the payoff from being the first. And in the second example, Chassang and Takahashi (2011) analyze a related model of the repeated Prisoners' Dilemma. Here the sucker payoff from being defected upon matters for whether cooperation can be sustained, not just the temptation payoff from being the defector. Furthermore, in Chassang and Takahashi's model, grim-trigger strategies need not be the best way to sustain cooperation: selective punishment strategies that punish defectors while attempting to reward cooperators sustain higher levels of cooperation by reducing the overall cost of being defected upon.

Finally, cooperation can also collapse when there is private monitoring, such as in the subjective-evaluation agency models by Levin (2003), MacLeod (2003), Fuchs (2007), and Maestri (2012). A key insight of this literature is that there must be inefficiency on the equilibrium path: to provide incentives, the agent must do worse when outcomes are poor; but to induce truthful revelation of outcomes, the principal must be indifferent across alternative output reports. See Malcomson (this volume) for more on private monitoring in relational incentive contracts.

6. The Clarity Problem

In Sections 5.2 and 5.3 we explored the credibility problem in relational contracting: should one party believe another's promise? In this section we finally reach the clarity problem in relational contracting: can one party understand another's promise? Both of these problems arise naturally if formal contracts are infeasible or imperfect because the parties are unable to articulate ex ante or to verify ex post their roles in and rewards from cooperating together. So far, the relational-contracts literature has focused on the credibility problem that arises when roles or rewards cannot be verified ex post. Here we begin to explore the possibility that roles or rewards cannot be fully articulated ex ante.

As an example familiar to many, consider the tenure criteria in an academic department. In principle, a department could use solely formal criteria (here meaning objective weights on objective measures); for example, department policy could be "You get tenure if and only if you have at least three papers in *Econometrica* by the fall of your sixth year." More typically, however, a department's tenure process allows some role for judgment and discretion—perhaps for reasons similar to those that led Lincoln Electric to complement its objective piece rate with a discretionary bonus, as described in Section 4.2. At Lincoln, decades of shared experience seem to have given managers and workers a shared sense for how the bonus program is supposed to be run (although see Section 6.1.2 for an interesting twist in this story). Likewise, in many departments, senior faculty seem to have a shared sense for what their department's tenure criteria are. The issue we explore in this section, however, is the possibility that these senior faculty cannot easily communicate their shared sense to a new assistant professor, or even to an outside senior hire. Clearly, phrases such as "have an impact on your field," "publish good papers in good journals," or "contribute appropriately to the work of the department" are only a beginning.

As a second example, recall from Section 4.3 the relational contract that supports propublication behavior by scientists and managers at science-driven pharmaceutical firms like Merck. Notice how much shared information is required. First, there is extensive task knowledge about the state-dependent actions that constitute propublication behavior. Each scientist has to learn not only what kinds of knowledge are potentially useful in the search for new drugs, but also how to behave almost like an academic scientist—including subtle expertise about which

conferences to go to, what kinds of papers to publish and, most importantly, how to make tradeoffs between behaving like an academic versus actively engaging in drug discovery. Likewise, each manager has to learn what kinds of rewards and recognition matter to research scientists—employees who famously care more about the design and reception of their work than they do about monetary rewards. All this learning, by scientists and managers, is significantly complicated by the fact that appropriate behaviors will likely vary across fields (e.g., cardiology versus neurology), across disciplines (e.g., chemists versus biologists), and with the particular circumstances of the firm (e.g., how strong are the firm's market and patent positions, how close is it to finding other new drugs, and so on).

In short, as Winter and others studying the inspiration problem described in Section 5.1 have emphasized, task knowledge is often both extensive and difficult to communicate. We further emphasize that management practices relying on relational contracts have additional information requirements, which we call relational knowledge. Roughly, if task knowledge concerns what is supposed to happen, relational knowledge is its complement: what will happen if parties don't do what they are supposed to—i.e., in game-theoretic terminology, what is supposed to happen off the equilibrium path? We note that, hard as it might be to communicate task knowledge, there is a sense in which relational knowledge is more difficult to communicate, because events off the equilibrium path are not supposed to happen, so there may be less opportunity for the parties to learn from experience about relational knowledge than about task knowledge.

Given the shared task and relational knowledge underlying propublication behaviors by scientists and managers, we find it impressive that Merck and some other pharmaceutical firms managed to succeed at this management practice. We also find it unsurprising that other firms were slower or less successful at implementing this practice; see Cockburn, Henderson, and Stern (2000). What we ask readers to imagine here, however, is how Merck got this practice going in its early days. We find it easy, for example, to imagine a recruiter trying to explain the novel approach to a newly minted post doc (who, let's say, would otherwise be considering academic jobs) by appealing to a rough metaphor such as "It's almost like being an assistant professor." We return to both of these ideas—difficulties in communication and the related reliance on metaphors—in Section 6.2.

Motivated by these examples of tenure and propublication behaviors, we proceed in two steps in the body of this section. First, in Section 6.1, we provide brief descriptions of case studies that can be interpreted as examples where the parties may not have had a shared understanding of their intended relational contract. Then, in Section 6.2, we discuss recent evidence and theory about how parties might build a shared understanding. We emphasize, however, that both the evidence and the theory in Section 6.2 are for the case where the parties have identical interests. In this sense, this work explores the resolution of the clarity problem in its pure form, without any complications from the simultaneous presence of the credibility problem that arises from imperfectly aligned interests. Exploring the interaction between these two problems is a leading issue for further research, both theoretical and empirical: if one party does something unexpected, does the other attribute it to miscommunication or gaming?

6.1 Case Studies of Imperfectly Shared Understandings

In this section we discuss the possibility that parties may not have a shared understanding of their roles in and rewards from cooperating together. In Section 6.1.1 we consider examples where the possibility of misunderstanding surfaced relatively early in the relationship. In Section 6.1.2 we turn to examples that arose after decades of a successful relationship.[19]

6.1.1 Imperfectly Shared Understandings Early in Relationships

Empirically, it seems unsurprising that parties may not have a shared understanding early in their relationship. Theoretically, however, we know of no models exploring how this possibility affects relational contracting. That is, even in relational-contract models with learning such as those discussed in Section 5, one party may be disappointed at what the other turns out to have known all along, but the parties are never surprised to find that they have misunderstood each other. We return to this issue in Section 6.2.3.

We use two case studies to illustrate the possibility of imperfectly shared understandings relatively early in relationships. Naturally, these case studies also admit other interpretations. In particular, both of the case studies we describe here can be interpreted as cousins of the types

[19] This section draws on Gibbons (2010) with permission from the *Annual Review of Economics*, Volume 2, © 2010 by Annual Reviews, http://www.annualreviews.org.

models in Section 5.2, such as Halac's (2012) model where the principal has private information about her value from continuing the relationship.

First, consider Stewart's (1993) account of how Credit Suisse, a large European bank, bought the US investment bank First Boston, taking the company private under the name CS First Boston (CSFB). At the time of the deal, there was much speculation about how the new firm would handle the inevitable differences in culture and firm policies, especially with regards to bankers' compensation. Although we cannot know the details of managers' attempts to identify and resolve these differences, the events that unfolded in the years immediately following the change in ownership reveal a lack of shared understanding about the determination of the bankers' annual bonus payments.

In the first two years that Credit Suisse controlled CSFB, all firms in the investment-banking industry performed poorly and CSFB bankers received bonuses that were lower than the historical average but comparable to bonuses paid at other investment banks. In the third year, however, CSFB performed better than previously and yet worse than its competitors, and Credit Suisse paid bonuses that were above bonuses at CSFB in the first two years but lower than those at other firms in the third year. A crisis ensued. In brief, CSFB bankers asserted that the bonus policy in their industry was to match the market, meaning that bonuses should be competitive with bonuses at other top-bracket firms. In contrast, the Swiss asserted that in *their* industry the bonus policy was pay for performance, meaning that a banker's bonus depended on how he and his bank performed. Note that these two policies make identical pay prescriptions when all firms in the industry have the same performance, as was broadly true in the first two years.

We make no attempt to adjudicate the parties' competing claims in this case. Nonetheless, we do draw two lessons from this case. First, it seems unlikely that the parties had common knowledge ex ante about the determinants of the bonus (since they could not then offer each other conflicting claims about what they had understood). Second, in expressing their claimed understandings, both parties use metaphors: match the market and pay for performance. That is, although one could imagine solely formal bonus plans receiving such labels (e.g., "Pay the same total bonus pool as at Goldman Sachs" or "Pay 30% of profits as bonuses"), such formal contracts were clearly not what the parties were saying they had understood to be in place.

As a second example, consider the Danish hearing-aid firm Oticon. In 1990 the firm launched radical empowerment of its product-development projects with a memo from CEO Lars Kolind titled "Think the Unthinkable" that envisioned project groups as akin to mini-businesses, each with its own resources, timeline, goals, and incentives. The initial results were strong and were attributed to the new organization; subsequent commentators used language like "bringing the market inside the firm" (see Foss (2003) for discussion).

We don't know whether Oticon's executives explicitly used the market metaphor to introduce the new organization, but suppose they did (or, suppose the metaphor was used by contemporaneous commentators). Oticon's project managers could then interpret the market metaphor as an initial approximation (akin to "almost like being an assistant professor") or as something closer to a literal promise to run the firm as an open market. If the former, then they would have expected the metaphor to be refined as events unfolded, clarifying just how much autonomy project managers actually would have in the new organization; if the latter, then they may have been surprised and upset (perhaps akin to the Wall Street bankers at CSFB) when the firm's Projects and Products Committee (staffed by the CEO and three senior managers) tightened control after the firm's portfolio of projects spiraled into disarray. The reaction of the firm's employees to the change suggests that the latter was the closer to the case.

Again, the possibility we take from this case is that the parties did not have a shared understanding ex ante. Perhaps no one thought that the market had literally been brought inside the firm, in the sense that project managers were now identical to entrepreneurs in their control over project decisions, and perhaps everyone understood that the old way of running the firm (before "Think the Unthinkable") no longer applied. But the gap between these extremes is very large, and we find it easy to imagine that project managers might have thought or hoped that the new way the firm was to be run would be closer to the former than the latter.

6.1.2 Imperfectly Shared Understandings in Decades-Old Relationships

In 1981, the cover of Johnson & Johnson's annual report read "Decentralization = Creativity = Productivity" (Aguilar and Bhambri, 1986: 1). For decades before and after, J&J was comprised of many disparate and nearly autonomous health-care businesses, ranging in size from a handful to thousands of employees. While the substantial freedom given to each business

was widely believed to increase innovation and initiative, it also made coordination more difficult.

This tension is by no means unique to J&J. For example, Alfred P. Sloan's (1963 (1990): 429-35) account of his years at General Motors is similar: "It has been a thesis of this book that good management rests on a reconciliation of centralization and decentralization, or "decentralization with co-ordinated control." … It must be apparent that co-ordinated decentralization is not an easy concept to apply. … The balance which is struck between corporate and divisional responsibility varies according to what is being decided, the circumstances of the time, past experience, and the temperaments and skills of the executives involved. … [T]he responsibility for determining administrative organization is a continuing one."

One example of this tension between initiative and coordination at J&J arose as hospitals found it increasingly burdensome to work directly with so many separate businesses, requesting that J&J instead create a single organization that could handle sales and distribution for all its relevant divisions. While J&J's competitors moved rapidly to this structure, the firm delayed making the change for several years, plausibly because it believed that having corporate headquarters create this new sales and distribution business would raise serious questions as to the nature of the relational contract between HQ and the existing businesses: if headquarters was prepared to limit these businesses' historical autonomy in this way now, what would happen in the future?

Our reading of this episode in J&J's history is that there was of course never any doubt about whether corporate headquarters *could* make this change in the organization structure, but there was concern about whether and how it *should*—in the sense that doing so might be interpreted by some as reneging on an understanding of how the tension between initiative and coordination was to be managed at J&J. Presumably, this episode concerning sales and distribution to hospitals was not the first time this tension surfaced at J&J, nor was it the last (for example, see Barrett (2003)). As Sloan put it, the "responsibility for determining administrative organization is a continuing one." Or, in our terms, even decades of shared experience may not fully clarify a crucial relational contract.

As a second example where decades of a successful relationship nonetheless left key decisions in doubt, we update the history of Lincoln Electric begun in Section 4.2. As Hastings (1999) describes, in the early 1990s rapid international expansion led to serious financial difficulties at Lincoln. For example, in 1992 the firm as a whole lost $46 million, even though operations in Cleveland had an excellent year. These losses forced Lincoln to consider how large a bonus pool to pay the Cleveland workers. (Recall that Lincoln's bonus is completely discretionary, not governed by contract or formula.) Broadly similar to CSFB, there were two possible views on the bonus Lincoln should pay in Cleveland: a large bonus based on Cleveland's excellent results, or a small or zero bonus based on results for the firm as a whole. Because Lincoln's overseas expansion was recent, this distinction between Cleveland profits and aggregate profits had never arisen.

In the event, Lincoln's managers decided to conform to the Cleveland workers' interpretation of the contract and paid $52 million in bonuses at the end of 1992, despite the fact that the firm had to take on debt of $250 million (63% of equity). Lincoln's financial difficulties continued into the following year while the Cleveland operation continued to excel. Lincoln's management team launched an intensive communication effort to implore the Cleveland workers to expand their efforts and hours even further in order to make up for the firm's aggregate losses and failed production efforts abroad and to "rescue" the company from violating covenant agreements. Workers responded by redoubling their efforts and raising the capacity utilization of the Cleveland plant from roughly 75% to nearly full capacity. This required current workers to work weekends and holidays and to give up vacation time. These efforts played a major role in the firm's being able to honor its loan agreements, and the firm paid out $55 million in bonuses at the year's end despite again having aggregate losses.

We interpret these events as suggesting that in some important respects the relational contract that had developed at Lincoln was incomplete. When the company operated only in Cleveland, the question of whether the bonus should reflect plant- or firm-wide profits was immaterial; it was only after the company was operating across the world that it became clear that management and employees might have different understandings of their relational contract. Under these conditions Lincoln's decision to honor their employees' understanding of the relational contract represented a substantial investment in strengthening their relationship—an

investment whose power was illustrated in the following year by the employees' continued commitment to the firm.

6.2 Towards Modeling the Clarity Problem

As noted above, this second approach to understanding implementation difficulties that might arise through relational contracts—based on the clarity problem rather than, say, path dependence—is much more speculative; in fact, we know of no existing models. On the other hand, we are mindful of Krugman's (1995: 27) dictum that "Like it or not, … [in economics] the influence of ideas that have not been embalmed in models soon decays." We therefore hazard the following avenues for future research.

6.2.1 Team-Theoretic Experiments and Models

There are some fascinating laboratory studies and models of parties trying to develop a shared understanding. To date, however, all of these analyses assume that the parties have identical interests. We discuss this work here, because it gives exciting hints about what might be done with imperfectly aligned interests, as in our settings of interest.[20]

In a pair of beautiful laboratory investigations, Weber and Camerer (2003) and Selten and Warglien (2007) study how parties learn to communicate. Weber and Camerer show one subject 16 pictures of people in office settings, and a light goes on next to one of the pictures. The other subject sees the same 16 pictures but in a different order and without the light. After the light goes on, the first subject picks up a phone and speaks to the second; each subject gets a payoff if the second subject can touch the correct picture in a specified amount of time. After the specified time has elapsed, the second subject learns which picture had the light next to it.

Selten and Warglien conduct a parallel exercise, but with a much simpler state space: instead of 16 pictures that are somewhat hard to describe, they use 6 simple figures, in a cross-product structure—a circle or a triangle on the outside of each figure, and nothing or a dot or a + on the inside. If the subjects could use a phone, the exercise would be too easy, because the first subject would simply say "Circle with a dot in it." But, having simplified the state space, Selten

[20] This section draws on Gibbons and Henderson, "Relational Contracts and Organizational Capabilities," *Organization Science*, by permission from the Institute for Operations Research and the Management Sciences, 7240 Parkway Drive, Suite 300, Hanover, MD 21076, USA, as well as on Gibbons (2010) with permission from the *Annual Review of Economics*, Volume 2, © 2010 by Annual Reviews, http://www.annualreviews.org.

and Warglien also simplify the message space, allowing the first subject to send characters from a small, abstract message space, such as {7, J, &, *, $}. The advantage of simplifying the message space in this way is that Selten and Warglien can easily capture the language used by the subjects over time. For example, the second subject might learn that 7J means a circle with a dot in it.

Formally, both of these papers watch players attempt to build a shared language during repetitions of a game like the following: (1) player 1 observes the state of the world, $s \in S$; (2) player 1 sends a costless message to player 2, $m \in M$; (3) player 2 chooses an action $a \in S$; (4) payoffs to each player are $U(s, a) = 1$ if $a = s$ and $U(s, a) = 0$ otherwise; (5) player 2 observes s. Thus, player 1 would like to send a message m(s) that tells player 2 that the state is s (so that player 2 will then choose the action $a = s$). The problem is that, at least in the early going, player 2 has little basis for understanding player 1's messages, especially in Selten and Warglien's setting.

Both papers find that different pairs of players develop different languages, even though these pairs are playing in the same environment (except for the random realizations of the states of the world over time). Because these are common-interest games, we interpret the knowledge being communicated as task knowledge, not relational knowledge. In particular, there is no concern with defection or punishment in a common-interest setting. Nonetheless, consistent with the large literature on the difficulties of communicating tacit knowledge, different pairs of players take different lengths of time to develop a shared understanding, and different pairs hold different shared understandings once they reach them. From this lab evidence, we find it only a small stretch to imagine that similar forces could cause reasonable people to hold different understandings in situations like those at Lincoln Electric or CSFB. Of course, the issue at these two companies involved bonuses, so these were not common-interest settings; rather, the problem of credibility also arose from imperfect alignment of interests.

Turning to theory, in a game with Pareto-ranked Nash equilibria and zero payoff to all players out of equilibrium, a focal point may command everyone's attention. But what if multiple Nash equilibria offer the best payoff, or some of the payoffs out of equilibrium are very negative (making coordination failure very costly)? Furthermore, what if the parties cannot easily discuss the opportunities they perceive (as when an organization has congealed into

functional silos and, say, the production and marketing groups have only a rudimentary language in common)?

Crawford and Haller (1990) provide a pioneering analysis of such issues in a repeated coordination game. One of their important insights is that, absent a common language about actions (e.g., about the detailed production and marketing activities that might be useful in concert), the parties' shared experience may facilitate coordination by allowing decentralized partners to label their action spaces in terms of past play.

Blume and Franco (2007) continue in this spirit, analyzing an n-player, m-action coordination game with k "successes" (Nash equilibria paying 1 to all players) and $m^n - k$ "failures" (action-tuples where all players receive 0). The parties know the number of successes but not the action-tuples that will achieve them. Each player observes his own actions and payoffs but not the actions of other players. The optimal strategy entails mixing (until a success is reached) so that the players do not all change their actions in lock-step (which would cause the players to revisit unsuccessful action-tuples). As a result of this mixing, different groups of n players could take different durations to find a success.

Whereas Blume and Franco call their work "Decentralized learning from failure," Ellison and Holden (2012) take a more hierarchical approach, in which a principal instructs an agent. Each period, (1) the agent observes the state of the world, $s \in S$; (2) the agent chooses an action $a \in A$; (3) the principal sends a message $m \in M$ to the agent; and (4) both parties receive the payoff $\pi(a,s)$. A novel aspect of the model is that the principal cannot communicate about a state until that state has been realized. More specifically, the principal's message dictates that if a future state is within a specified neighborhood of this period's state then the agent should take a specified action. When messages are of this form, there are more and less useful realizations of s that may occur in early periods; in particular, a useful realization is one that allows the principal to specify a broad neighborhood. As a result, dyads whose early realizations of s are useful will perform better.

Finally, as a first step away from common interests, there are a few papers that consider the costs that parties bear in trying to communicate with each other, focusing more on whether the parties will choose to incur these costs and what quality of communication will then occur, rather

than on what language is then used to communicate. The initial paper in this spirit is Dewatripont and Tirole (2005), who treat the problem as one of moral hazard in teams: will a sender and a receiver both invest sufficiently to allow an intelligible message to be sent by the one and understood by the other.

Closer to our concerns, Li (2012) studies a setting where the clarity problem is difficult to resolve within organizations because (1) propagation of tacit knowledge requires extensive shared experience between individuals, and thus can diffuse only gradually across the group, and (2) propagation of shared knowledge is privately costly, so self interest may stymie the propagation of knowledge. Initial conditions within an organization may then have a persistent effect on the extent to which knowledge is propagated across the organization and thus the extent to which the clarity problem is resolved for the organization as a whole. For example, if an organization expands too quickly for the diffusion of knowledge to keep up with the expansion, the resultant incomplete propagation may persist even after the period of rapid growth has ended.

6.2.2 Categories and Metaphors

We next step away from economics, as a first step toward seeing whether help might be available elsewhere. We find that it may very well be.

Celebrated work in cognitive psychology and linguistics has explored the nature of categories and metaphors. For example, Rosch (1973, 1975) departed from the view that categories are defined by a set of properties shared by all their members. Instead, Rosch showed that there can be more and less representative members of categories, ranging from prototypes to outliers. Lakoff and Johnson (1980) and Lakoff (1987) explored related issues concerning metaphors. See Mullainathan, Schwartzstein, and Shleifer (2007) for an initial economic model of some of these ideas.

The ensuing stream of research from cognitive psychology and linguistics has important implications for the difficulty in communicating relational contracts. After all, parties use relational contracts when formal contracts (say, with objective weights attached to objective measures) are imperfect or do not exist. In the early days of science-based drug discovery at Merck, for example, one could imagine a recruiter trying to explain the novel approach to a young scientist by appealing to a rough metaphor such as “It’s almost like being an assistant

professor." Similarly, we describe an example below of attempting to "Bring the market inside the firm."

The existence of sharper and fuzzier categories and metaphors suggests that it may be easier both to express and to learn a sharper metaphor (such as "Act like you own the company."), but this very sharpness may mean that the metaphor does a poorer job capturing what is intended and needed. Similarly, for a given category or metaphor, it may be that learning is faster and more reliable from exposure to prototypes rather than outliers, but the availability of such examples may be random over time. The prevalence of story-telling in organizations can be seen as an attempt to promulgate prototypical examples.

6.2.3 Potential Paths Forward

We close this section with a few very loose ideas about potential paths forward. One is illustrated by the case of CSFB, where the two sides ostensibly held quite different views about compensation policy (match the market and pay for performance) that nonetheless delivered identical pay prescriptions when all firms in the industry had the same performance, as was broadly true in the first two years after Credit Suisse bought First Boston. More generally, parties with different understandings may not appreciate that such a difference exists until key events occur—with the further complication that behavior by one party during such an event may be intended as cooperation but nonetheless be viewed by the other as defection. Loosely speaking, the CSFB case evokes for us something like Fudenberg and Levine's (1993) definition of self-confirming equilibrium (weaker than Nash because the players' understandings of each others' strategies need agree only on the equilibrium path); see also Ryall (2003) for applications to strategy. Relatedly, see Greif (1994) and Fudenberg and Levine (2003) for analyses of how very different equilibria can be if one group holds one belief about how play would proceed where nodes off the equilibrium path to be reached, but a second group holds a different belief. Greif, for example, interprets the family firm in Genoa versus community enforcement among the Maghribi traders in this way.

Finally, another appealing avenue is to take seriously the prospect that people may really misunderstand each other. A beautiful recent example is Ettinger and Jehiel's (2010) theory of deception, where parties have only coarse knowledge of each other's strategies. This approach

connects nicely with the well documented bias from social psychology, the fundamental attribution error (Ross, 1977), in which parties over-attribute another person's behavior to that person's permanent type, rather than to transitory shocks to that person's environment. See Repenning and Sterman (2002) for a system-dynamics argument about the role of such errors in the dynamics of process improvement. It would be outstanding if economic modeling like Ettinger-Jehiel were ultimately able to connect social psychology like Ross with system dynamics and process improvement like Repenning-Sterman.

7. Conclusion

In this chapter we have explored the role of relational contracts in sustaining persistent performance differences among seemingly similar enterprises. We began by reviewing the evidence that PPDs among SSEs exist and are economically significant across a wide range of industries and geographies, and we summarized the evidence that these performance differences are correlated with variation in management practices. Critically, we suggested that the successful performance of many of these practices cannot be easily articulated ex ante or verified ex post, and we used three qualitative accounts to suggest that many (most?) competitively significant management practices rely on the presence of relational contracts.

We then turned to the question of why managerial practices, if they do indeed play a significant role in enabling persistent performance differences, do not diffuse more readily. We reviewed the rich literature exploring this question and suggested that it can be complemented by a focus on the difficulties firms may encounter in building relational contracts. Here we described three barriers firms may face: bad parameters, bad luck, and bad communication. First, the relational contracts operated by leading firms may be either infeasible or prohibitively costly for underperformers to implement. In this case, measured PPDs are due to bad parameters (i.e., unmeasured heterogeneity in the benefits and/or costs of any particular relational contract). Second, relationships may be path dependent: the sequence of events experienced during the relationship can produce measured performance differences among ex ante identical enterprises. We suggested that achieving perfunctory cooperation can make it harder to achieve consummate cooperation; that cooperation, once built, can be fragile; and that cooperation may be difficult to build in the first place. Third and less conventionally, we suggested that the problem of clarity—

difficulty in communicating the extensive task and relational information that underlies many relational contracts—may also play a role in making it difficult to build unfamiliar relational contracts.

At our most ambitious, we hope that our specific discussion of managerial practices and PPDs will help shape a broader research agenda in organizational economics about what, exactly, managers do. For example, For example, existing models of heterogeneity in relational contracts due to bad parameters take various characteristics as given. New work endogenizing firm features could explore important managerial opportunities to increase the feasible set (and to reduce the costs) of building relational contracts—for example, through recruitment and acculturation of types, design of network structure, and so on. In a similar vein, existing models emphasizing the importance of path dependence take bad luck as given. Future analyses could explore ways in which managers may reduce the probability of perfunctory cooperation or protect fragile initial cooperation.

The third area—relating to problems of clarity—offers perhaps the broadest opportunity for new theoretical and empirical work. We hope our discussion of categories and codes as well as our examples of the ways in which clarity created problems provides inspiration for new research. A more complete treatment of managerial practices, for example will almost certainly need to examine the interaction of credibility and clarity.

In addition to new theory these issues cry out for empirical work focused sharply on relational contracting. Gibbons and Henderson (2012) give some illustrations of relational contracts in laboratory experiments and field data on contracts between firms, but there is perilously little analogous work inside organizations. One promising start, however, is Bloom, Sadun, and Van Reenen (QJE, forthcoming), who use the large-sample methodology of Bloom and Van Reenen described in Section 3, but complement it with auxiliary data on trust scores between regions. Bloom et al. find that a multinational with headquarters in one country will grant more decentralization (measured, say, by expenditure constraints) to a plant in a second country than it will to a plant in a third country if the trust score from country 1 to 2 is higher than from country 1 to 3. We see this as an enormously promising start toward empirical work on the issues we have raised here but clearly much more remains to be done.

Finally, moving from our main focus on productivity to the neighboring notion of profitability, much of what we have described can be cast in reduced form as saying that relational contracts are an investment that might improve an enterprise's productivity. But investments are costly, and there are typically many such investments a firm could make and one intriguing question is therefore whether and how investments in relational contracts might act as substitutes or complements to other more conventional assets. Both Toyota and Southwest airlines, for example, appear to have used investment in relational contracts as routes to enter industries that had for many years been dominated by firms with harder assets—superior brands and prime geographic locations. We suspect that exploring the interaction between relational contracts and other forms of investment might well be a productive target for future research.

REFERENCES

Abreu, Dilip. 1988. "On the Theory of Infinitely Repeated Games with Discounting." *Econometrica* 56: 383–396.

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica* 58: 1041–1063.

Adler, Paul, and Charles Heckscher. 2006. "Towards Collaborative Community," in Charles Heckscher and Paul Adler (eds.), *The Firm as a Collaborative Community.* New York: Oxford University Press, pp. 11–105.

Aguilar, Francis, and Arvind Bhambri. 1986. "Johnson & Johnson (B): Hospital Services." Harvard Business School Case 384-054. Boston: Harvard Business School Press.

Ali, Nageeb, and Navin Kartik. 2012. "Herding with Collective Preferences." *Economic Theory,* forthcoming.

Almeida, Paul, and Bruce Kogut. 1999. "Localization of Knowledge and the Mobility of Engineers in Regional Networks." *Management Science* 45: 905–917.

Anand, Bharat, and Tarun Khanna. 2000. "Do Firms Learn to Create Value? The Case of Alliances." *Strategic Management Journal* 21(special issue: Strategic Networks): 295–315.

Angrist, Josh, Parag Pathak, and Christopher Walters. 2012. "Explaining Charter School Effectiveness." NBER Working Paper 17332, National Bureau of Economic Research, Cambridge, MA.

Appelbaum, Eileen, Thomas Bailey, Peter Berg, and Arne Kalleberg. 2000. *Manufacturing Advantage: Why High-Performance Work Systems Pay Off.* Ithaca, NY: Cornell University Press.

Appleyard, Melissa, Nile Hatch, David Mowery. 2000. "Managing the Development and Transfer of Process Technologies in the Semiconductor Manufacturing Industry," in Giovanni Dosi, Richard Nelson, and Sidney Winter (eds.), *The Nature and Dynamics of Organizational Capabilities.* London: Oxford University Press, pp. 183–207.

Arrow, Kenneth. 1974. *The Limits of Organization.* New York: W. W. Norton.

Arthur, Jeffrey. 1994. "Effects of Human Resource Systems on Manufacturing Performance and Turnover." *Academy of Management Journal* 37: 670–687.

Bailey, Martin, Charles Hulten, David Campbell, Timothy Bresnahan, and Richard Caves. 1992. "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity: Microeconomics* (1992): 187–267

Bar-Isaac, Heski. 2003. "Reputation and Survival: Learning in a Dynamic Signaling Model." *Review of Economic Studies* 70: 231–251.

Barnard, Chester. 1938. *The Functions of the Executive.* Cambridge, MA: Harvard University Press.

Barney, Jay. 1986. "Organizational Culture: Can It Be a Source of Sustained Competitive Advantage?" *Academy of Management Review* 11: 656–665.

Barrett, Amy. 2003. "Staying On Top." *Business Week,* May 5, pp. 60–68.

Bartelsman, Eric, and Mark Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature* 38: 569–594.

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

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *Quarterly Journal of Economics* 118: 1169–1208.

Biorn, Eric, Kjersti-Gro Lindquist, and Terje Skjerpen. 2002. "Heterogeneity in the Returns to Scale: A Random Coefficient Analysis with Unbalanced Panel Data." *Journal of Productivity Analysis* 18: 39–57.

Black, Sandra, and Lisa Lynch. 2001. "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity." *Review of Economics and Statistics* 83: 434–445.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122: 1351–1408.

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

———. 2011. "Human Resource Management and Productivity," in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics,* volume 4. Amsterdam: Elsevier and North-Holland, pp. 1697–1767.

Bloom, Nicholas, Carol Propper, Stephen Seiler, and John Van Reenen. 2010. "The Impact of Com- petition on Management Quality: Evidence from Public Hospitals." NBER Working Paper 16032, National Bureau of Economic Research, Cambridge, MA.

Bloom, Nicholas, Benn Eifert, David McKenzie, Aprajit Mahajan, and John Roberts. 2012a. "Does Management Matter? Evidence from India." Working paper, Stanford University, Stanford, CA.

Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen. 2012b. "Management Practices across Firms and Countries." *Academy of Management Perspectives* 26: 12–33.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012c. "Americans Do I.T. Better. US Multinationals and the Productivity Miracle." *American Economic Review* 102: 167–201.

———. "The Organization of Firms across Countries." 2012d. *Quarterly Journal of Economics,* forth- coming.

———. 2012e. "Management as Technology." Working paper, Stanford University, Stanford, CA. Blume, Andreas, and April Franco. 2007. "Decentralized Learning from Failure." *Journal of Economic Theory* 133: 504–523.

Bonatti, Alessandro, and Johannes Ho¨rner. 2011. "Collaborating." *American Economic Review* 101: 632–663.

Boning, Brent, Casey Ichniowski, and Kathryn Shaw. 2007. "Opportunity Counts: Teams and the Effectiveness of Production Incentives." *Journal of Labor Economics* 25: 613–650.

Breschi, Stefano, and Francesco Lissoni. 2009. "Mobility of Inventors and Networks of Collaboration: An Anatomy of Localised Knowledge Flows." *Journal of Economic Geography* 9: 439–468.

Bresnahan, Timothy, Erik Brynjolfsson, and Loren Hitt. 2002. "Information Technology, Workplace Organization and the Demand for Skilled Labor: Firm-level Evidence." *Quarterly Journal of Economics* 117: 339–376.

Bresnahan, Timothy, Shane Greenstein, and Rebecca Henderson. 2011. "Schumpeterian Competition and Diseconomies of Scope: Illustrations from the Histories of Microsoft and IBM," in Josh Lerner and Scott Stern (eds.), *The Rate and Direction of Inventive Activity, 50th Anniversary Volume.* Cambridge, MA: National Bureau of Economic Research, pp. 203–271.

Bull, Clive. 1987. "The Existence of Self-Enforcing Implicit Contracts." *Quarterly Journal of Economics* 102: 147–159.

Callander, Steven. 2011. "Searching and Learning by Trial and Error." *American Economic Review* 101: 2277–2308.

Chassang, Sylvain. 2010a. "Fear of Miscoordination and the Robustness of Cooperation in Dynamic Global Games with Exit." *Econometrica* 78: 973–1006.

———. 2010b. "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Rela- tional Contracts." *American Economic Review* 100: 448–465.

Chassang, Sylvain, and Gerard Padro´ i Miquel. 2010. "Conflict and Deterrence under Strategic Risk." *Quarterly Journal of Economics* 125: 1821–1858.

Chassang, Sylvain, and Satoru Takahashi. 2011. "Robustness to Incomplete Information in Repeated Games." *Theoretical Economics* 6: 49–93.

Chew, Bruce, Kim Clark, and Timothy Bresnahan. 1990. "Measurement, Coordination and Learning in a Multiplant Network," in Robert Kaplan (ed.) *Measures for Manufacturing Excellence.* Boston: Harvard Business School Press, pp. 129–162.

Christensen, Clayton. 1997. *The Innovator's Dilemma.* Boston: Harvard Business School Press.

Clark, Kim, and Takahiro Fujimoto. 1991. *Product Development Performance: Strategy, Organization, and Management in the World Auto Industry.* Boston: Harvard Business School Press.

Cockburn, Iain, Rebecca Henderson, and Scott Stern. 2000. "Untangling the Origins of Competitive Advantage." *Strategic Management Journal* 21(special issue): 1123–1145.

Cohen, Wesley, and Daniel Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35: 128–152.

Crawford, Vincent, and Hans Haller. 1990. "Learning How to Cooperate: Optimal Play in Repeated Coordination Games." *Econometrica* 58: 571–595.

Cyert, Richard, and James March. 1963. *A Behavioral Theory of the Firm.* Oxford: Blackwell.

Disney, Richard, Jonathan Haskel, and Ylva Heden. 2003. "Restructuring and Productivity Growth in UK Manufacturing." *Economic Journal* 113(July): 666–694.

Dobbie, Will, and Roland G. Fryer. 2012. "Getting beneath the Veil of Effective Schools: Evidence from New York City." Working paper, Harvard University, Cambridge, MA.

Dunlop, John, and David Weil. 1996. "Diffusion and Performance of Modular Production in the U.S. Apparel Industry." *Industrial Relations* 35: 334–355.

Dwyer, Douglas.1998. "Technology Locks, Creative Destruction, and Nonconvergence in Productivity Levels." *Review of Economic Dynamics* 1: 430–437.

Ellison, Glenn, and Richard Holden. 2012. "A Theory of Rule Development." *Journal of Law, Economics, and Organization,* forthcoming.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler. 2004. "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Columbia." *Journal of Development Economics* 75: 333–371.

Ettinger, David, and Philippe Jehiel. 2010. "A Theory of Deception." *American Economic Journal: Microeconomics* 2: 1–20.

Fainmesser, Itay. 2012. "Community Structure and Market Outcomes: A Repeated Games-in-Networks Approach." *American Economic Journal: Microeconomics* 4: 32–69.

Fast, Norman, and Norman Berg. 1975. *The Lincoln Electric Company.* Harvard Business School Case 9-376-028. Boston: Harvard Business School Press.

Foss, Nicolai. 2003. "Selective Intervention and Internal Hybrids: Interpreting and Learning from the Rise and Decline of the Oticon Spaghetti Organization." *Organization Science* 14: 331–349.

Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2006. "Market Selection, Reallocation, and Restruc- turing in the U.S. Retail Trade Sector in the 1990s." *Review of Economics and Statistics* 88: 748–758.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98: 394–425.
Friedman, James. 1971. "A Non-Cooperative Equilibrium for Supergames." *Review of Economic Studies* 38: 1–12.

Fuchs, William. 2007. "Contracting with Repeated Moral Hazard and Private Evaluations." *American Economic Review* 97: 1432–1448.

Fudenberg, Drew, and David Levine. 1993. "Self-Confirming Equilibrium." *Econometrica* 61: 523–545.

———. 2006. "Superstition and Rational Learning." *American Economic Review* 96: 630–651.

Gant, Jon, Casey Ichniowski, and Kathryn Shaw. 2002. "Social Capital and Organizational Change in High-Involvement and Traditional Work Organizations." *Journal of Economics and Management Strategy* 11: 289–328.

Gibbons, Robert. 2010. "Inside Organizations: Pricing, Politics, and Path Dependence." *Annual Review of Economics* 2: 337–365.

Gibbons, Robert, and Rebecca Henderson. 2012. "Relational Contracts and Organizational Capabilities." *Organization Science,* forthcoming.

Gibbons, Robert, and Michael Waldman. 1999. "Careers in Organizations: Theory and Evidence," in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics,* volumes III and IV. New York: North-Holland, pp. 2373–2437.

Gittell, Jody Hoffer. 2003. *The Southwest Airlines Way: Using the Power of Relationships to Achieve High Performance.* New York: McGraw-Hill.

Greif, Avner. 1994. "Cultural Beliefs and the Organization of Society: A Historial and Theoretical Reflection on Collectivist and Individualist Societies." *Journal of Political Economy* 102: 912–950.

Griliches, Zvi. 1986. "Productivity, R&D, and Basic Research at the Firm Level in the 1970s." *American Economic Review* 76: 141–154.

Griliches, Zvi, and Jacques Mairesse. 1981. "Productivity and R&D at the Firm Level." NBER Working Paper 826, National Bureau of Economic Research, Cambridge, MA.

———. 1982. "Comparing Productivity Growth: An Exploration of French and US Industrial and Firm Data." NBER Working Paper 961, National Bureau of Economic Research, Cambridge, MA.

———. 1985. "R&D and Productivity Growth: Comparing Japanese and U.S. Manufacturing Firms." NBER Working Paper 1778, National Bureau of Economic Research, Cambridge, MA.

Halac, Marina. 2012. "Relational Contracts and the Value of Relationships." *American Economic Review* 102: 750–779.

Hamilton, Barton, Jack Nickerson, and Hideo Owan. 2003. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy* 111: 465–497.

Hastings, Donald. 1999. "Lincoln Electric's Harsh Lessons from International Expansion." *Harvard Business Review* 77: 162–178.

Hawanini, Gabriel, Venkat Subramanian, and Paul Verdin. 2003. "Is Performance Driven by Industry or Firm-Specific Factors? A New Look at the Evidence." *Strategic Management Journal* 24: 1–16.

Henderson, Rebecca. 1994. "The Evolution of Integrative Capability: Innovation in Cardiovascular Drug Discovery." *Industrial and Corporate Change* 3: 607–630.

Henderson, Rebecca, and Kim Clark. 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms." *Administrative Science Quarterly* 35: 9–30.

Henderson, Rebecca, and Iain Cockburn. 1994. "Measuring Competence? Exploring Firm Effects in Pharmaceutical Research." *Strategic Management Journal* 15(S1): 63–84.

Holmes, Thomas, David Levine, and James Schmitz. 2012. "Monopoly and the Incentive to Innovate When Adoption Involves Switchover Disruption." *American Economic Journal: Microeconomics* 4: 1–33.

Hoxby, Caroline, and Sonali Murarka. 2009. "Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement." NBER Working Paper 14852, National Bureau of Economic Research, Cambridge MA.

Hsieh, Chang-Tai, and Peter Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124: 1403–1448.

Huselid, Mark. 1995. "The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance." *Academy of Management Journal* 38: 635–672.

Huselid, Mark, and Brian Becker. 1996. "Methodological Issues in Cross-Sectional and Panel Estimates of the Human Resource-Firm Performance Link." *Industrial Relations* 35: 400–422.

Iansiti, Marco, and Kim Clark. 1994. "Integration and Dynamic Capability: Evidence from Product Development in Automobiles and Mainframe Computers." *Industrial and Corporate Change* 3: 557–605.

Ichniowski, Casey, and Kathryn Shaw. 1999. "The Effects of Human Resource Systems on Productivity: An International Comparison of U.S. and Japanese Plants." *Management Science* 45: 704–722.

Ichniowski, Casey, Kathryn Shaw, and Giovanni Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *American Economic Review* 87: 291–313.

Jackson, Matthew, Tomas Rodriguez-Barraquer, and Xu Tan. 2012. "Social Capital and Social Quilts: Network Patterns of Favor Exchange." *American Economic Review,* forthcoming.

Jeitschko, Thomas, and Curtis Taylor. 2001. "Local Discouragement and Global Collapse: A Theory of Coordination Avalanches." *American Economic Review* 91: 208–224.

Kaplan, Sarah. 2011. "Strategy and PowerPoint: The Epistemic Culture and Machinery of Strategy Making." *Organization Science* 22: 320–346.

Kaplan, Sarah, Fiona Murray, and Rebecca Henderson. 2003. "Discontinuities and Senior Management: Assessing the Role of Recognition in Pharmaceutical Firm Response to Biotechnology." *Industrial and Corporate Change* 12: 203–233.

Klette, Tor. 1999. "Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data." *Journal of Industrial Economics* 47: 451–476.

Kochan, Thomas, Harry Katz, and Robert McKersie. 1986. *The Transformation of American Industrial Relations.* New York: Basic Books.

Kreps, David. 1990. "Corporate Culture and Economic Theory," in James Alt and Kenneth Shepsle (eds.), *Perspectives on Positive Political Economy.* New York: Cambridge University Press, pp. 90– 143.

———. 1996. "Markets and Hierarchies and (Mathematical) Economic Theory." *Industrial and Cor- porate Change* 5: 561–595.

Krugman, Paul. 1995. *Development, Geography, and Economic Theory.* Cambridge, MA: MIT Press.

Lacetera, Nicola, Iain M. Cockburn, and Rebecca Henderson. 2004. "Do Firms Change Capabilities by Hiring New People? A Study of the Adoption of Science-Based Drug Discovery." *Advances in Strategic Management* 21: 133–159.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lazear, Edward. 2000. "Performance Pay and Productivity." *American Economic Review* 90: 1346–1361.

Leibenstein, Harvey. 1969. "Organizational or Frictional Equilibria, X-Efficiency, and the Rate of Innovation." *Quarterly Journal of Economics* 83: 600–623.

———. 1987. *Inside the Firm: The Inefficiencies of Hierarchy.* Cambridge, MA: Harvard University Press.

Levin, Jonathan. 2003. "Relational Incentive Contracts." *American Economic Review* 93: 835–857.

———. 2009. "The Dynamics of Collective Reputation." *B.E. Journal of Theoretical Economics* 9(1): Article 27.

Levinthal, David. 1997. "Adaptation on Rugged Landscapes." *Management Science* 43: 934–950.

Li, Jin, and Niko Matouschek. 2011. "Managing Conflicts in Relational Contracts." Mimeo, Northwestern University, Evanston, IL.

Liker, Jeffrey K. 2004. *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer.* New York: McGraw-Hill.

Liker, Jeffrey K., and James Franz, 2011. *The Toyota Way to Continuous Improvement.* New York: McGraw-Hill.

Liker, Jeffrey K., and David Meier. 2006. *The Toyota Way Fieldbook.* New York: McGraw-Hill.

Lipitz-Snyderman, Allison, Donald Steinwachs, Dale Needham, Elisabeth Colantuoni, Laura Morlock, and Peter Pronovost. 2011. "Impact of a Statewide Intensive Care Unit Quality Improvement Initiative on Hospital Mortality and Length of Stay: Retrospective Comparative Analysis." *BMJ* 342: d219.

Lippert, Steffen, and Giancarlo Spagnolo. 2011. "Networks of Relations and Word-of-Mouth Communication." *Games and Economic Behavior* 72(1): 202–217.

Liu, Qingmin. 2011. "Information Acquisition and Reputation Dynamics." *Review of Economic Studies* 78: 1400–1425.

Lucas, Robert. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22: 3–42.

MacDuffie, John Paul. 1995. "Human Resource Bundles and Manufacturing Performance: Organizational Logic and Flexible Production Systems in the World Auto Industry." *Industrial and Labor Relations Review* 48: 197–221.

MacDuffie, John Paul, Kannan Sethuranman, and Marshall Fisher. 1996. "Product Variety and Manufacturing Performance: Evidence from the International Automotive Assembly Plant Study." *Management Science* 42: 350–369.

Macher, Jeffrey. 2006. "Technological Development and the Boundaries of the Firm: A Knowledge- Based Examination in Semiconductor Manufacturing." *Management Science* 52: 826–843.

Macher, Jeffrey, and David Mowery. 2003. "Managing Learning by Doing: An Empirical Study in Semiconductor Manufacturing." *Journal of Product Innovation Management* 20: 391–410.

———. 2009. "Measuring Dynamic Capabilities: Practices and Performance in Semiconductor Manufacturing." *British Journal of Management* 20: S41–S62.

MacLeod, Bentley. 2003. "Optimal Contracting with Subjective Evaluation." *American Economic Review* 93: 216–240.

MacLeod, Bentley, and James Malcomson. 1988. "Reputation and Hierarchy in Dynamic Models of Employment." *Journal of Political Economy* 96: 832–854.

———. 1989. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica* 57: 447–480.

Maestri, Lucas. 2012. "Bonus Payments versus Efficiency Wages in the Repeated Principal-Agent Model with Subjective Evaluations." *American Economic Journal: Microeconomics* 4: 34–56.

Malcomson, James. 2012. "Relational Incentive Contracts with Persistent Private Information." Working paper, University of Oxford.

McAdams, David. 2011. "Performance and Turnover in a Stochastic Partnership." *American Economic Journal: Microeconomics* 3: 107–142.

McGahan, Anita, and Michael Porter. 1997. "How Much Does Industry Matter, Really?" *Strategic Management Journal* 18 (summer special issue: Organizational and Competitive Interactions): 15– 30.

———. 1999. "The Persistence of Shocks to Profitability." *Review of Economics and Statistics* 81: 143– 153.

Melitz, Marc. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 17: 1695–1725.

Milgrom, Paul, and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organizations." *American Economic Review* 80: 511–528.

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

Mintzberg, Henry. 2004. *Managers not MBAs: A Hard Look at the Soft Practice of Managing and Management Development.* San Francisco: Berrett-Koehler.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. "Coarse Thinking and Persuasion." *Quarterly Journal of Economics* 123: 577–619.

Mundlak, Yair. 1961. "Empirical Production Function Free of Management Bias." *Journal of Farm Economics* 43: 44–56.

Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change.* Cambridge, MA: Harvard University Press.

Ohno, Taiichi. 1988. *Toyota Production System: Beyond Large-Scale Production.* Portland, OR: Productivity Press.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64: 1263–1297.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* New York: Cambridge University Press.

Penrose, Edith. 1959. *The Theory and Growth of the Firm.* New York: John Wiley and Sons.

Pfeffer, Jeffrey. 1998. *The Human Equation: Building Profits by Putting People First.* Boston: Harvard Business School Press.

Pil, Frits, and John MacDuffie. 1996. "The Adoption of High-Involvement Work Practices." *Industrial Relations* 35: 423–455.

Powell, Michael. 2012. "Productivity and Credibility in Industry Equilibrium." Working paper, Northwestern University, Evanston, IL.

Pronovost, Peter, and Eric Vohr. 2010. *Safe Patients, Smart Hospitals.* New York: Hudson Street Press. Pronovost Peter, Dale Needham, Sean Berenholtz, David Sinopoli, Haitao Chu, Sara Cosgrove, Bryan Sexton, Robert Hyzy, Robert Welsh, Gary Roth, Joseph Bander, John Kepros, and Christine Goeschel. 2006. "An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU." *New England Journal of Medicine* 355: 2725–2732.

Pronovost Peter, Christine Goeschel, Elizabeth Colantuoni, Sam Watson, Lisa Lubomski, Sean Beren- holtz, David Thompson, David Sinopoli, Sara Cosgrove, Bryan Sexton, Jill Marstellar, Robert Hyzy, Robert Welsh, Patricia Posa, Kathy Schumacher, and Dale Needham. 2010. "Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study." *BMJ* 340: c309.

Reinganum, Jennifer. 1989. "The Timing of Innovation: Research, Development and Diffusion," in Richard Schmalensee and Robert Willig (eds.), *Handbook of Industrial Organization,* volume 1. Amsterdam: North-Holland, pp. 849–908.

Repenning, Nelson, and John Sterman. 2002. "Capability Traps and Self-Confirming Attribution Errors in the Dynamics of Process Improvement." *Administrative Science Quarterly* 47: 265–295.

Rivkin, Jan. 2000. "Imitation of Complex Strategies." *Management Science* 46: 824–844.

Rivkin, Jan, and Nicolaj Siggelkow. 2003. "Balancing Search and Stability: Interdependencies among Elements of Organizational Design." *Management Science* 49: 290–311.

Rosch, Eleanor. 1973. "Natural Categories." *Cognitive Psychology* 4: 328–350.

——. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104(3): 192–233.

Ross, Lee. 1977. "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process," in Leonard Berkowitz (ed.), *Advances in Experimental Social Psychology,* volume 10. New York: Academic Press, pp. 173–220.

Rousseau, Denise. 1995. *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements.* Thousand Oaks, CA: SAGE.

Rumelt, Richard. 1991. "How Much Does Industry Matter?" *Strategic Management Journal* 12: 167–185.

Ryall, Michael. 2003. "Subjective Rationality, Self-Confirming Equilibrium, and Corporate Strategy." *Management Science* 49: 936–949.

Salter, Wilfred E. G. 1960. *Productivity and Technical Change.* Cambridge: Cambridge University Press.

Schein, Edgar. 1985. *Organizational Culture and Leadership.* San Francisco: Jossey-Bass.
Schmalensee, Richard. 1985. "Do Markets Differ Much?" *American Economic Review* 75: 341–351.

Selten, Reinhard, and Massimo Warglien. 2007. "The Emergence of Simple Languages in an Experimental Coordination Game." *Proceedings of the National Academy of Sciences (USA)* 104: 7361– 7366.

Siggelkow, Nicolaj, and Jan Rivkin. 2005. "Speed and Search: Designing Organizations for Turbulence and Complexity." *Organization Science* 16: 101–122.

Simon, Herbert. 1947. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization.* New York: Free Press.

Sloan, Alfred P. 1963. *My Years with General Motors.* New York: Doubleday. Reprinted in 1990 by Crown Publishing Group, New York.

Smith, Lones, and Peter Sørensen. 2000. "Pathological Outcomes of Observational Learning." *Econometrica* 68: 371–398.

Spector, Robert, and Patrick McCarthy. 2012. *The Nordstrom Way to Customer Service,* second edition. Hoboken, NJ: Wiley.

Stewart, James. 1993. "Taking the Dare." *New Yorker,* July 26, pp. 34–39.

Syverson, Chad. 2004a. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 86: 534–550.

———. 2004b. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy* 112: 1181–1222.

———. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49: 326–365.

Tirole, Jean. 1996. "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)." *Review of Economic Studies* 63: 1–22.

Tripsas, Mary, and Giovanni Gavetti. 2000. "Capabilities, Cognition, and Inertia: Evidence from Digital Imaging." *Strategic Management Journal* 21: 1147–1161.

Van Biesebroeck, Johannes. 2008. "The Sensitivity of Productivity Estimates: Revisiting Three Important Debates." *Journal of Business and Economic Statistics* 26: 311–328.

Walker, Francis. 1887. "The Source of Business Profits." *Quarterly Journal of Economics* 1: 265–288.

Watson, Joel. 1999. "Starting Small and Renegotiation." *Journal of Economic Theory* 85: 52–90.
———. 2002. "Starting Small and Commitment." *Games and Economic Behavior* 38(1): 176–199.

Weber, Roberto, and Colin Camerer. 2003. "Culture Conflict and Merger Failure: An Experimental Approach." *Management Science* 49: 400–415.

Winter, Sidney. 1987. "Knowledge and Competence as Strategic Assets," in David Teece (ed.), *The Competitive Challenge: Strategies for Industrial Innovation and Renewal.* Cambridge, MA: Ballinger, pp. 159–184.

———. 1988. "On Coase, Competence, and the Corporation." *Journal of Law, Economics, and Organization* 4: 163–180.

———. 2006. "Toward a Neo-Schumpeterian Theory of the Firm." *Industrial and Corporate Change* 15: 125–141.

Wolitzky, Alex. 2012. "Cooperation with Network Monitoring." *Review of Economic Studies,* forthcoming.

Womack, James, Daniel Jones, and Daniel Roos. 1991. *The Machine That Changed the World: The Story of Lean Production.* New York: HarperCollins.

Yang, Huanxing. 2012. "Nonstationary Relational Contracts with Adverse Selection." *International Economic Review,* forthcoming.

Zollo, Maurizio, and Sidney Winter. 2002. "Deliberate Learning and the Evolution of Dynamic Capabilities." *Organization Science* 13: 339–351.