

Firm-Wide Incentives and Mutual Monitoring at Continental Airlines

Marc Knez, *Lexecon Strategy Group*

Duncan Simester, *Massachusetts Institute of Technology*

In February 1995 Continental Airlines introduced an incentive scheme that promised monthly bonuses to all 35,000 hourly employees if the company achieved a firm-wide performance goal. Conventional wisdom suggests that free riding will render such schemes ineffective. We present evidence indicating that the incentive scheme raised employee performance despite the apparent threat of free riding. To explain why the scheme may have been effective we argue that the organization of employees into autonomous work groups enabled Continental to induce mutual monitoring among employees within each work group.

I. Introduction

Firm-wide incentive schemes are common at large firms. A 1987 Government Accounting Office survey of 326 Fortune 1,000 firms reported that 54% of nonunion and 39% of unionized firms had profit-sharing schemes for hourly employees (cited in Cooke 1994). Moreover, their use

We wish to thank Eric Anderson, Michael Browne, Colin Camerer, Rob Gertner, John Hauser, and Birger Wernerfelt for their helpful comments. The article has benefited from comments by seminar participants at the University of Chicago, Columbia University, MIT, Stanford University, and the Boston/Cambridge Seminar on the Economics of Organizations. Many people at Continental Airlines also contributed to this study. For their generous contributions of time and data we would particularly like to thank Gordon Bethune, Gregory Brenneman, Beth Ballew, Thomas Yourke, and the employees and supervisors at Midway Airport, Chicago, Illinois.

[*Journal of Labor Economics*, 2001, vol. 19, no. 4]
© 2001 by The University of Chicago. All rights reserved.
0734-360X/2001/1904-0001\$02.50

appears to be increasing, both in the form of employee stock ownership plans (ESOP) and profit-sharing plans (Conte and Kruse 1991). This prevalence contrasts sharply with accepted economic wisdom that employee free riding will render such schemes ineffective. We investigate the apparent inconsistency between managerial practice and common wisdom by studying the impact of a firm-wide incentive scheme implemented at one large firm (Continental Airlines).

In February 1995, following a sustained period of poor operational and financial performance, Continental Airlines introduced an incentive scheme that promised a monthly bonus to every hourly employee if the airline met firm-wide performance goals. We first investigate claims that the scheme contributed to improvements in Continental's performance. The data offer support for these claims, and so we investigate the factors that may have mitigated free riding.

Evaluating the Impact of Continental's Incentive Scheme

Previous empirical work includes several studies that investigate how incentives affect decisions and behavior in small groups. These studies generally reveal modest positive associations between incentives and group performance; however, the results may not generalize beyond small groups.¹ Previous studies of firm-wide incentive schemes at large firms include substantial literatures investigating both ESOPs and profit-sharing schemes. In a review of 27 studies of ESOPs, Blasi, Conte, and Kruse (1996) conclude that few of the findings reveal strong and statistically significant effects of employee ownership on performance. Studies of profit sharing have tended to yield slightly more positive outcomes. Kruse (1993) identifies a total of 26 studies of profit sharing, representing data from several countries. Within these studies he found 265 reported coefficients measuring the impact of profit sharing, and reports that just 8.7% of these coefficients take negative values, while 57.4% take positive values that are significantly different from zero ($p < .05$). The results of Kruse's own analysis conducted using a new data source also reveal a positive association between profit sharing and average productivity increases.

With few exceptions, these ESOP and profit-sharing studies compare performance across firms and so are hampered by the multitude of unobserved idiosyncratic factors that influence firm performance. Other studies compare the performance of a single firm over time but cannot control for intervening industry or firm changes that may also have affected performance. Resolving both types of limitations generally requires a cross-sectional comparison of time-series performance, preferably within the

¹ See, e.g., Cooke (1994) and Welbourne, Balkin, and Gomez-Mejia (1993) and the references cited therein.

same firm. Studying the impact of firm-wide incentives requires a comparison of performance improvements across similar business units, where some units participate in the incentive program and others do not. The data we use from Continental offer a rare example of these conditions in a field setting.

In Section II we describe the group incentive scheme introduced by Continental and attempt to isolate the impact of the scheme from the effects of other changes. We do so by comparing the improvement in on-time departure performance at two groups of airports. Airports in the first (test) group are all locations at which Continental's operations were performed by its own employees. In the second (control) group these operations were performed by employees of separate companies with which Continental had outsourcing contracts. Because only Continental employees were eligible for the bonus scheme, the bonus scheme was not expected to affect behavior at airports where operations were outsourced. Outsourced airports therefore provide us with a natural control group to (cross-sectionally) separate the impact of the bonus scheme from intervening firm or industry changes. Moreover, by comparing performance of individual airports over time, we control for airport differences that may also have affected performance.

Our analysis reveals a significantly larger improvement at nonoutsourced airports. This result is consistent with the bonus scheme raising the efforts of employees that were eligible to receive it and survives several tests designed to rule out alternative interpretations. To explain why the incentive scheme may have been effective, we build on recent work in both economics and sociology that suggests that mutual monitoring may provide an effective solution to free riding.

Free Riding and Mutual Monitoring

Compensating employees for improvements in joint output leads to employees sharing the rewards from higher effort. The dilution in the resulting incentives and mitigation of additional effort is often termed "free riding" and is aggravated in large firms.² In small firms, employees share their rewards with fewer colleagues, while mutual monitoring may also help to resolve free riding. In particular, if employees collusively

²This observation forms the basis of Alchian and Demsetz's (1972) theory distinguishing partnerships and firms. They argue that the weakening of incentives in partnerships leads to a transfer of residual rights to a single principal who enforces efficient effort by direct monitoring. It also provides the starting point for Holmström's (1982) work on moral hazard in teams. Holmström formalizes the incentives to free ride and shows that the problem is exacerbated in large firms. He concludes that overcoming the free-rider problem may yield an alternative role for the principal: enforcing a system of budget-breaking penalties (or bonuses) in order to implement efficient effort.

agree to exert high effort and then monitor and sanction their colleagues to enforce the group decision, incentives to exert effort will increase. Group incentive schemes may encourage monitoring and sanctioning because each employee's actions affect payments to other members of the group.

Unfortunately, mutual monitoring is not expected to be effective in large firms. Employees in large groups are often unable to observe each other's efforts and are less willing to incur the costs of monitoring and sanctioning their colleagues (Heckathorn 1988; Coleman 1990; Kandel and Lazear 1992). Both concerns are relevant at Continental. The bonus scheme includes approximately 35,000 eligible employees, none of whom can directly affect joint output. Moreover, employees are dispersed among different terminal and airport locations, which prevents direct monitoring of each other's efforts. Notwithstanding these obstacles, our explanation for the claimed success of Continental's incentive scheme argues that Continental exploits the benefits of mutual monitoring, despite its size. It does so because employees are organized into autonomous groups, dispersed between airports and terminals. Within these groups, collusion is sustainable. Moreover, the interdependence of the groups amplifies the impact of poor performance, so that firm-level measures are sensitive to individual group performance. As a result, rewards for improving firm-level performance are sufficient to prompt choice of high effort by each group.

The remainder of the article is in two parts, beginning with a review of the impact of the incentive scheme on Continental's performance, followed by an investigation of the factors that may have mitigated the apparent free riding problem. The empirical findings in Section II begin with a description of the incentive scheme introduced at Continental. We then attempt to isolate the impact of the scheme from the effects of other changes. In Section III we explain why we might expect mutual monitoring (collusion) to fail at a firm of Continental's size. We then argue that collusion may succeed within Continental's autonomous work groups and present anecdotal evidence consistent with this claim. Next we describe how the actions of each work group affect overall firm output. Finally, we discuss the possible existence of multiple equilibria and how improving the flight schedule may have helped move to a high effort equilibrium. Section IV contains a summary of conclusions and limitations.

II. Continental Airline's Incentive Scheme

Prior to 1995, Continental Airlines was consistently one of the worst-performing airlines in the industry. Since deregulation of the industry in 1978 it had entered bankruptcy protection twice, failed to make an annual

profit during periods it was not under bankruptcy protection, and averaged last among the 10 major domestic airlines in on-time arrival, baggage handling, and customer complaints.³ At the end of 1994, a new senior management team was brought in to address these problems. This new team represented the tenth change in senior management in 10 years and was greeted with an immediate cash crisis. Persistent (and apparently accurate) rumors were circulating that the company might not be able to meet its January 1995 payroll obligations.

The new team unveiled a multifaceted turnaround strategy labeled the "Go Forward Plan." Three actions were taken to improve on-time arrival and departure performance: changing airport managers, improving the flight schedule, and introducing a group incentive scheme that paid a monthly bonus if a firm-wide on-time performance goal was met.⁴ The new flight schedule was implemented on January 9, 1995, and was designed to make on-time arrivals and departures feasible by reducing schedule conflicts and increasing turnaround times (slack) for equipment and personnel. The changes in airport managers involved replacing or rotating the senior manager at every airport, with the changes occurring throughout 1995 and 1996.

The bonus scheme was announced on January 15, 1995, and promised \$65 to every hourly employee (including part-time employees) in every month that Continental's on-time performance ranked among the top five in the industry. The rankings were based on the proportion of flights arriving on time (within 15 minutes of schedule) as reported by the Department of Transportation (DOT). In months when the bonus was paid, checks for the full \$65 were distributed by managers to the approximately 35,000 eligible employees, at a cost of approximately \$2.3 million (additional taxes for the bonus were deducted from regular payroll payments). The program began in February 1995 and continued unchanged until December of that year. In 1996, the scheme was modified, paying \$65 in months in which Continental ranked second or third (in on-time arrival) and \$100 when it finished first.⁵

There are several features of on-time arrival that made it an attractive performance measure. First, on-time arrival was accurately measured and

³ This assessment is based on Continental's 1994 and 1995 annual reports and the Department of Transportation's Air Travel Consumer Reports, which report monthly on-time arrival, baggage handling, and customer complaint statistics for the 10 major domestic airlines.

⁴ Other elements of the Go Forward Plan are described in more detail in the company's annual reports.

⁵ It is interesting that this change was planned when the scheme was first introduced. After describing the new bonus plan in the 1994 annual report, the CEO predicted that the performance goal would increase as performance improved.

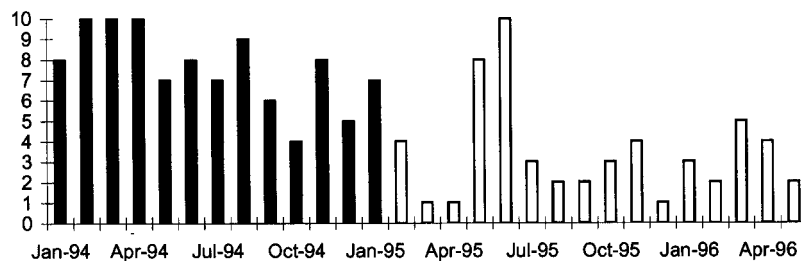


FIG. 1.—Department of Transportation on-time performance rankings: Continental's rank among the 10 major airlines

could easily be compared with that of competitors. Arrival and departure times were recorded automatically when airplanes entered and left gates, and this information was collected from each airline and reported by the DOT. Second, on-time performance was directly affected by the actions of the air crew and airport employees.⁶ Finally, it was widely believed that failing to arrive on time led to other difficulties, including delayed baggage and dissatisfied customers.⁷

By the end of 1996 the threat of bankruptcy was averted. After reporting net losses of \$125 million in 1992, \$199 million in 1993, and \$613 million in 1994, Continental reported a net profit of \$224 million in 1995. This grew to \$319 million in 1996 and \$385 million in 1997, and was accompanied by large operational cash flow surpluses. The changes directed at improving on-time performance were also successful. The airline ranked in the top half of the DOT on-time performance rankings in 9 of the 11 months following introduction of the bonus scheme in February 1995 (see fig. 1). The 2 months in which the goal was not achieved (May and June) coincided with disruptive industrial action by the airline's pilots. Similar improvements were observed in the delayed baggage and customer complaint statistics.

Continental's senior management attributed some of the improvement in Continental's on-time performance to the success of the bonus scheme. Their confidence in the effectiveness of the scheme stemmed from observed changes in employee behavior. In addition to anecdotes illustrating

⁶ We describe how employees' actions affected on-time performance in more detail in a later section. At least one of the competing airlines (Southwest) recorded and reported arrival and departure times manually.

⁷ To verify this wisdom we collected the DOT's monthly performance data describing late arrivals, lost baggage, and customer complaints for the top 10 airlines for the years 1989-95 and found very significant correlations between all three measures ($p < .01$). Although the correlations between the baggage and complaint measure disappeared when on-time performance was partialled out, the correlations with late arrivals survived partialing out the third variable.

an increase in employee effort and mutual monitoring of co-workers, they cited reductions in employee turnover, on-the-job injuries, and sick days. When we asked about the possibility that the improvements were due to other changes, we were reminded that the bonus program was the primary activity directed at improving employee performance. At the suggestion of the senior managers we interviewed employees and/or airport managers at several Continental airports. Employees and airport managers almost universally volunteered the bonus scheme as one of the reasons for the turnaround in Continental's on-time performance.⁸ Many employees believed that the success of the scheme was self-evident. After recognizing that it had affected their behavior, employees at one airport argued for introduction of a similar scheme to improve performance elsewhere in the company.

The incentive scheme was also self-funding. After introduction of the scheme fewer Continental customers missed connections and had to be reaccommodated on other airlines, and other airlines used Continental more frequently to reaccommodate their own customers. This yielded additional cash flow of over \$8 million per month, yet the cost of the incentive scheme was less than \$3 million per month (Brenneman 1998).

In the remainder of this section, we attempt to isolate the impact of the firm-wide bonus scheme to determine whether the scheme did in fact contribute to the performance improvements. We do so by exploiting Continental's use of outsourcing at several airports. Because only Continental employees were eligible for the bonus scheme, the bonus was not expected to affect behavior where activities were outsourced. Hence, if the bonus scheme was effective we would expect smaller performance improvements at airports that were outsourced. In contrast, the changes in both the flight schedule and airport management were system-wide, designed to improve performance at every airport (Continental had managers at all airports, including locations at which activities were outsourced).

If the bonus scheme was effective, we might have expected introduction of similar on-time performance bonuses in outsourcing contracts. Discussions with Continental management revealed that these contracts typically did include bonuses for achieving on-time performance goals. However, there was no systematic change in these bonuses during the course of the study. Moreover, several contracting difficulties limited the size of performance bonuses in outsourcing contracts. First, due to interdependencies between flights, on-time performance at one airport was affected by the behavior of employees at other airports. Second, if the compensation received by outsourcing partners depended largely on performance at their airport, the resulting focus on performance at that airport may

⁸ We describe the interview process in detail in a later section.

have had a detrimental impact on the rest of the system.⁹ Finally, controlling for environmental factors by comparing performance with the rest of the industry was difficult at an airport level. Individual airports tended to be serviced primarily by two to three carriers, so that scale differences make comparisons difficult.

All of our data (except for local weather conditions) were provided by Continental Airlines. We begin by describing our dependent measure and the nature and extent of outsourcing at the 32 domestic airports in the study. These 32 airports include Continental's three hubs, its busiest non-hub domestic destinations and a sample of smaller locations (the 32 airports represent a large portion of Continental's domestic route structure). To control for differences between airports we introduce a series of control variables and compare the improvements in on-time performance under these controls. We finish this section by reviewing alternative explanations for the results.

Dependent Measure: Change in On-Time Departure Performance

Our primary dependent measure is the change in on-time departure performance at each airport. Focusing on change in performance allows us to control for the numerous unobservable (exogenous) factors that affected performance at each airport. The decision to use on-time departure rather than arrival was made after discussions with airport managers revealed that departure measures more accurately reflect individual airport performance. On-time arrival performance is subject to delays on departure or in flight, which are largely outside the control of the destination airport.

A departure was considered on time if it left the gate within 15 minutes of the scheduled departure time (a standard industry measure). For each of the 32 airports in the study, Continental provided us with monthly data describing the proportion of flights that departed on time for the period January 1994 through November 1996. The measures are extremely accurate (they are recorded electronically when planes depart from the gates) and summarize performance for an average of approximately 1,000 monthly flights at each airport. We focus on the first month that employees could receive the bonus (February 1995) through the end of the data period (November 1996). Our dependent measure (Improvement) was calculated as the proportion of flights that departed on time in the

⁹ Analogous reasoning explains why payment of the performance bonus was not extended to employees of the outsourcing partners. This may have led to these employees focusing on Continental flights to the detriment of other flights that they service.

Table 1
Summary of Outsourcing Activity

Airport	Type of Outsourcing	Date Outsourcing Was Introduced
Las Vegas	Full	October 1994
Phoenix	Full	October 1994
Seattle	Full	April 1995
Salt Lake City	Full	February 1995
Washington, D.C.	Partial	Before 1994
Denver	Partial	November 1994
Detroit	Partial	April 1995
Orlando	Partial	Before 1994
San Diego	Partial	Before 1994
Tampa	Partial	Before 1994

month minus the percentage that departed on time in the corresponding month in 1994.¹⁰

Outsourcing

The independent variable of interest is the presence or absence of outsourcing. The activities at each airport were divided into two main categories: ramp and gate operations. The ramp operations included guiding in and pushing out aircraft and loading and unloading baggage. Gate operations comprised ticketing and checking in passengers, updating passengers' documentation and reservations, answering passenger queries, positioning the air-bridge, and enplaning and deplaning passengers. In addition there were some incidental activities that were performed by separate (often outsourced) employees, including fueling, catering, cleaning, and maintenance. We received data describing whether gate and/or ramp activities were outsourced at each airport.

Outsourcing of gate and/or ramp activities to other airlines or ground-handling firms occurred at 10 of the 32 airports in the sample (see table 1). We distinguished between full and partial outsourcing using two dichotomous variables: Full Outsourcing and Partial Outsourcing. In months in which an airport was fully outsourced, the Full Outsourcing variable was set equal to one, and in other months the variable was set equal to zero. The Partial Outsourcing variable was constructed in an analogous manner.

Discussions with airport managers revealed that the introduction of outsourcing typically leads to a 2–3-month period of upheaval while the

¹⁰ For example, the dependent measure for July 1995 equals the proportion of flights that departed on time in July 1995 minus the proportion of flights that departed on time in July 1994. The dependent measure for July 1996 equals the proportion of flights that departed on time in July 1996 minus the proportion of flights that departed on time in July 1994.

outsourced employees learned how to integrate Continental's flights into their other operations. These short-term distortions could increase or decrease Improvement (the dependent variable) depending on whether outsourcing was introduced in 1994 or 1995. To ensure that we did not misattribute any short-term effects of outsourcing we constructed a variable to identify the 3 months following the introduction of outsourcing and allowed this variable to take values of one, zero, and negative one. If outsourcing was introduced less than 3 months before the current month, the variable took a value of one; if it was introduced in 1994 within 3 months of the corresponding (comparison) 1994 month, it took a value of negative one; otherwise it was zero. Two separate variables were calculated in this manner to describe the introduction of full and partial outsourcing. Later in this section we discuss the factors that led to outsourcing at some airports and the impact that the introduction of outsourcing had on on-time performance.

Control Variables

To control for possible systematic variance between outsourced and nonoutsourced airports we collected several measures of airport differences. Our first concern was to identify factors that determined which airports were outsourced. Discussions with the Chief Operating Officer (COO), who was responsible for the introduction of additional outsourcing in 1994 and 1995, identified cost as the only criterion. Infrequent flights lead to equipment (tugs, luggage carts, de-icing trucks) lying idle and create difficulties in scheduling employees. As a result, ramp operations can often be performed at lower cost by an airline or ground-handling firm that also services other flights at the same airport. Economies of scale are not as large in gate operations where equipment (computer terminals) is specific to each airline and has to be supplied irrespective of outsourcing.¹¹

We collected four measures to control for scale differences between airports. First, the variable Continental's Scale describes the monthly number of departures from each airport in each month (in thousands). The second scale variable, Continental's Presence, reflects the proportion of an airport's flights that were operated by Continental and is intended to control for the availability of scale economies through outsourcing. The variable was calculated by dividing the number of monthly Continental departures by the total number of departures across all airlines at each airport for that month. Because it takes longer to load passengers and baggage when flights are full we calculated an approximate load factor each month by dividing the number of passengers by the number of flights

¹¹ Increased control of customer interactions was also suggested as an explanation for why gate operations are less likely to be outsourced.

(Load Factor). Finally, changes in the scale of Continental's operations at any airport may have affected performance either because greater complexity made coordination more difficult or because employees were harder to monitor. To control for changes in scale we calculated the average percentage change in the number of monthly departures (Change in Continental's Scale) from the corresponding month in 1994.¹²

We were concerned that poor on-time performance in 1994 may have contributed to the decision to outsource some airports. This was denied by management, who claimed that outsourcing was determined solely by cost factors. Nevertheless, to control for this issue we calculated a measure of each airport's 1994 on-time performance (1994 Performance). This measure represented the percentage of flights departing on time for the corresponding month in 1994. Including 1994 Performance as an independent variable also accounts for possible regression to the mean due to stochastic factors not captured by the other variables. Moreover, it explicitly recognizes the possibility that it may be harder to improve performance when initial performance is high (we also address this issue with additional analysis in later discussion).

The literature on organizational change suggests that resistance to change increases with employee tenure. We included two variables describing the average service time of nonmanagement employees at each airport, where the two variables distinguished between part-time and full-time employees (Part-Time Service and Full-Time Service). This distinction was prompted by discussions with airport managers, who observed that part-time employees at partially outsourced airports were often employees who had full-time positions prior to outsourcing and sought to return to that status. These employees generally had longer service records and were more highly motivated than other part-time employees who did not have career ambitions with Continental. Service time measured years employed in any capacity at Continental as of November 1996 (the date for which we had data). Because service time was not available for outsourced employees, for the four fully outsourced airports we used the mean for the other 28 airports.¹³

To control for the possibility that sustained periods of poor weather may have affected performance at some airports, we included a measure of weather differences (Weather) between each month and the corresponding month in 1994. The data were provided by the National Climatic

¹² We also considered hub effects and nonlinear scale effects. The data did not support these effects and so, for the sake of brevity, we omit them from the discussion.

¹³ Setting these variables to either the mean of the 22 nonoutsourced airports or the mean of the six partially outsourced airports (for the four fully outsourced airports) made little difference to the findings.

Data Center (NCDC) and measured monthly precipitation (in hundreds of inches) at each airport.

If flight schedule changes favored the nonoutsourced airports we would expect to see larger performance improvements at these airports. Other than a possible focus on the hub cities (which would be reflected in our attempts to identify scale effects) we have no reason to expect that flight schedule changes benefited the nonoutsourced airports more than the outsourced airports. However, the change in on-time arrival at each airport may offer a way to test and control for this possibility. The ramp and gate activities occurred after a flight arrived and before it departed. As a result, although schedule changes were expected to influence both arrival and departure performance, on-time arrival was less likely to be influenced by ramp and gate activities than on-time departure. Hence, improvements in on-time arrival offer a measure of schedule improvements that is somewhat independent of any change in the behavior of ramp and gate employees at that airport.¹⁴ The Schedule Improvement variable measures improvement in on-time arrival by subtracting average on-time arrival performance from performance in the corresponding month in 1994. A flight is considered on time if it arrived within 15 minutes of the scheduled arrival time, and (like departures) these data are recorded electronically and instantly transmitted to Continental's Houston operations center.

Continental employed an airport manager at all of the airports in the sample, including the partially and fully outsourced airports. At all but four airports the airport managers were changed during the measurement period. It is possible that these management changes may have affected the outsourced and nonoutsourced airports differently, either because one group of airports received the better managers or because the role of managers varies at outsourced airports. To address these possibilities the variable New Manager was set equal to one in months after a management change and zero otherwise. We then separately interacted New Manager with Full Outsourcing and Partial Outsourcing to allow the effects of a management change to vary across outsourced and nonoutsourced airports.

There are two airports in our sample for which operations were closed down and then later restarted during our measurement period. One of these airports was fully outsourced when it reopened and the other was not outsourced. Although we were unable to get detailed information on

¹⁴ To the extent that arrival performance also reflects behavior changes due to the on-time bonus, inclusion of this control will (conservatively) lead to underestimation of the impact of the bonus. To support the claim that ramp and gate employees have less influence over on-time arrival performance than departure performance we used the outsourcing and control variables to predict the improvement in on-time arrivals. As we expected, the outsourcing coefficients were not significant (see later discussion).

Table 2
Descriptive Statistics

Variables	Mean	Standard Deviation
Improvement	5.19	7.04
Full Outsourcing	.12	.33
Partial Outsourcing	.19	.39
3 Months after Full Outsourcing	.01	.16
3 Months after Partial Outsourcing	.00	.10
Continental's Scale	.93	1.83
Continental's Presence	.12	.20
Load Factor	80.14	21.87
Change in Continental's Scale	-.17	.36
1994 Performance	83.02	7.19
Part-Time Service	6.04	2.92
Full-Time Service	10.37	3.81
Weather	300.41	270.99
Schedule Improvement (Arrivals)		
Improvement	4.78	10.33
New Manager	.49	.50
New Manager with Full Outsourcing	.05	.22
New Manager with Partial Outsourcing	.09	.28
Reopened	.03	.18

NOTE.—The sample size for each variable is 648 (the sample used in the multivariate analysis).

changes that occurred at these two airports, the changes were apparently more wide-ranging than at other airports. The variable Reopened was set equal to one in months after these two airports had been reopened and zero otherwise.

Results

Summary statistics for all of the variables are presented in table 2 (sample sizes vary due to missing data, caused largely by airport closures). A simple comparison of means suggests that the level of outsourcing did moderate the improvement in on-time performance. The average improvement at nonoutsourced airports was 5.6% (from 82.4% to 88.0%), which compares with an improvement of just 4.9% at partially outsourced airports and 2.8% at fully outsourced airports. The difference between the nonoutsourced and fully outsourced airports is statistically significant ($5.6 > 2.8$, $p < .01$), as is the difference between the partially and fully outsourced airports ($4.9 > 2.8$, $p < .05$).

Although we omitted months immediately following the introduction of outsourcing from these means, this comparison does not control for scale, weather, or any of the other sources of systematic variation between outsourced and nonoutsourced airports that were expected to affect the rate of improvement. To control for the effects of the pilots' strike and other fixed monthly effects we included dummy variables separately identifying the 22 months in the sample (February 1995 through November

Table 3
Results of Multivariate Analysis

Variables	Full Model	Basic Model
Full Outsourcing	-3.42 (.66)	-1.52 (.81)
Partial Outsourcing	-1.63 (.58)	-.37 (.46)
3 Months after Full Outsourcing	-4.34 (1.13)	-7.31 (2.08)
3 Months after Partial Outsourcing	-2.53 (1.12)	-7.69 (2.59)
Continental's Scale	-.53 (.23)	
Continental's Presence	4.57 (2.08)	
Load Factor	-.03 (.01)	
Change in Continental's Scale	-2.11 (.85)	
1994 Performance	-.55 (.03)	
Part-Time Service	.41 (.09)	
Full-Time Service	-.13 (.06)	
Weather	-.31 (.06)	
Schedule Improvement (Arrivals Improvement)	.20 (.03)	
New Manager	-1.60 (.42)	
New Manager with Full Outsourcing	2.16 (1.07)	
New Manager with Partial Outsourcing	1.39 (.66)	
Reopened	4.91 (1.69)	

NOTE.—Standard errors are in parentheses. The dependent variable is Improvement. The sample size and adjusted R^2 are 680 and .39 for the basic model and 648 and .76 for the full model.

1996).¹⁵ Given the panel nature of the data and the use of 1994 data as a common benchmark, we used generalized least squares (GLS) to correct the standard errors. The results are presented in table 3, omitting estimates of the fixed monthly effects. As a basis for comparison we also present a much simpler version of the model, in which we control only for the short-term effects of introducing outsourcing together with the fixed monthly effects.

In the full model the coefficients for Full Outsourcing and Partial Outsourcing are significantly less than zero ($p < .01$) and indicate that improvements at the fully and partially outsourced airports were significantly smaller than at the nonoutsourced airports. The relative magnitudes of the two coefficients suggests that improvements diminished with the level of outsourcing (the difference in the magnitude of the two coefficients is significant, $p < .05$). These findings could be interpreted as evidence that the bonus scheme improved the performance only of those employees eligible to receive it. If so, the findings offer support for claims by Continental's employees and senior management that the bonus program did have an impact on employee performance. The outsourcing

¹⁵ We do not include fixed airport effects because our outsourcing variables are largely invariant across time. This suggests that it may be appropriate to aggregate the time-series data and conduct a cross-sectional comparison of improvement in on-time performance across the 32 airports. The results when data are aggregated in this manner are described in later discussion.

coefficients are also both negative in the basic model. We caution that interpreting the coefficients in this model is difficult, as the omission of relevant variables is likely to have biased the coefficients.

The other coefficients in the full model are also of interest, at least to the extent that they accord with their predicted effects:

Outsourcing appeared to have a particularly detrimental impact on performance in the 3 months immediately following its introduction.

High load factors, many departures (a large scale), and an increase in scale all acted to restrict improvements. Improvements appeared to be larger at airports at which Continental had a greater presence.

The negative coefficient for 1994 Performance is consistent with both regression to the mean and fewer opportunities to improve when initial performance is already high. The accuracy of the performance measures favors the initial performance interpretation, although regression to the mean may result from other stochastic factors.

The employee service coefficients suggest that resistance to change increases with the tenure of full-time employees. In contrast, the rate of improvement increases with the tenure of part-time employees.

Deterioration in the weather does appear to lead to flight delays.

When on-time arrival performance improved, on-time departure performance also improved.

Management changes appear to have had a more favorable effect at airports with more outsourcing. This could be due to reassignment of the better managers to the outsourced airports or could reflect the different role of managers at outsourced airports.

There were larger improvements in on-time performance following the temporary closures of the two reopened airports.

In response to a reviewer's concern that the arrivals measure may have biased the other coefficient estimates, we reestimated the model after omitting this measure. There was very little change in the coefficients for the remaining variables. We also reestimated the full model using the improvement in arrivals as the dependent measure. Neither of the outsourcing coefficients were significant ($p < .10$), which is consistent with our interpretation that ramp and gate crews at the arriving airport have less influence on the timing of arrivals than departures. We also further investigated the two service measures by including a variable describing the proportion of part-time employees at each airport. The coefficient was not close to significance, and inclusion of this variable had little effect on the other coefficients.

Although the results in table 3 support the claim that the bonus program had a direct impact on employee performance, the limited temporal variance in the two outsourcing variables may lead to concern as to whether a pooled time-series approach is appropriate. Recognizing this concern, we considered an alternative approach under which the data for each

Table 4
Results from the Aggregate Model

Variables	Basic Model	Full Model	Stepwise Model
Full Outsourcing	-2.72 (1.88)	-4.79 (2.77)	-3.41 (1.32)
Partial Outsourcing	-.37 (1.59)	-2.07 (2.76)	-2.14 (1.29)
Continental's Scale		-.41 (1.06)	
Continental's Presence		8.02 (10.10)	4.29 (2.27)
Load Factor		-.03 (.03)	
Change in Continental's Scale		-6.64 (1.93)	-6.96 (1.72)
1994 Performance		-.20 (.11)	-.17 (.10)
Part-Time Service		.58 (.27)	.53 (.23)
Full-Time Service		-.22 (.19)	-.24 (.16)
Weather		.92 (1.86)	
Schedule Improvement (Arrivals Improvement)		.29 (.15)	.31 (.18)
New Manager		-2.14 (1.65)	-1.68 (1.12)
New Manager with Full Outsourcing		3.75 (3.68)	
New Manager with Partial Outsourcing		.32 (2.82)	
Reopened		4.09 (2.55)	5.34 (2.03)

NOTE.—Standard errors are in parentheses. The dependent variable is Improvement. The sample size for all three models is 32. The adjusted R^2 is .003 for the basic model, .54 for the full model, and .61 for the stepwise model.

airport are aggregated and a comparison is made between the outsourced and nonoutsourced airports.

Aggregate Model

We compare on-time performance in the 9 months between January 1996 and September 1996 with on-time performance between January 1994 and September 1994. These time periods exclude the pilot's strike, most of the management changes, and the introduction of outsourcing in 1995 and at the end of 1994. The dependent variable was aggregated to represent the difference at each airport in average on-time performance between the period January 1996–September 1996 and the period January 1994–September 1994. An analogous approach was used to average the various scale measures, 1994 Performance, Weather, Schedule Improvement, and the management change variables. The remaining variables did not vary over the respective aggregation periods. We estimated three models, and the coefficients for these models are presented in table 4. The first model includes just the two outsourcing variables, and the second model includes all of the relevant control variables. In the third model, we used stepwise regression to select variables.

Both outsourcing coefficients variables entered the stepwise model. The outsourcing coefficients all maintained their negative signs, and the relative magnitudes are generally consistent with the corresponding models in

table 3 (the larger standard errors are consistent with the loss of statistical power due to aggregation). It appears that despite the conservative nature of this approach the evidence that on-time performance improvements were smaller at the outsourced airports is robust. Reassuringly, the coefficients for the other variables estimated in the full model are also generally consistent with those reported in table 3. We turn next to a discussion of alternative explanations for the outsourcing results.

Introduction of Outsourcing

Outsourcing was introduced during our measurement period at all four fully outsourced airports and at two partially outsourced airports. Our results suggest that the introduction of outsourcing led to short-term disruptions in performance. If the negative effects of outsourcing were more enduring, the introduction of outsourcing could explain the smaller observed improvements in performance where it was introduced. There are reasons to believe that this was not the case.

Focusing on the two airports at which partial outsourcing was introduced (Denver and Detroit) with the four airports at which it was already present reveals a larger improvement where it was introduced. Between February 1995 and November 1996 the average improvement in on-time departure at Denver and Detroit was 6.2%, compared to an average improvement of just 4.4% at Washington, D.C., Orlando, San Diego, and Tampa. If the introduction of outsourcing had worsened performance, we would expect to have seen smaller (not greater) performance increases at these two airports.

We also reestimated the models in table 3 after separating the Partial Outsourcing variable into two variables denoting partial outsourcing prior to 1994 and partial outsourcing in 1994 and 1995. In the full model these coefficients were both negative, and in the basic model the coefficients were consistent with the average improvements at Denver and Detroit and the other four partially outsourced airports (reported above). In particular, in the reestimated basic model the partial outsourcing coefficient for Denver and Detroit was positive and the coefficient for the other four airports was negative. If the introduction of outsourcing had worsened performance we would have again expected smaller improvements at Denver and Detroit than at the four airports that were partially outsourced before 1994.¹⁶

¹⁶ Furthermore, the average 1994 on-time performance at the four airports outsourced prior to 1994 was higher than the average of the other 28 airports. If outsourcing had a negative effect on performance we would expect lower (not higher) 1994 performance at the four airports that were partially outsourced in 1994.

Table 5
Initial Performance and Management Changes

Variables	Column 1	Column 2
Full Outsourcing	-3.74 (.65)	-3.66 (.60)
Partial Outsourcing	-2.11 (.58)	-1.92 (.69)
3 Months after Full Outsourcing	-4.54 (1.05)	-4.68 (1.04)
3 Months after Partial Outsourcing	-1.75 (1.22)	-4.34 (1.47)
Continental's Scale	-.28 (.21)	.15 (.28)
Continental's Presence	2.78 (1.99)	-.78 (2.68)
Load Factor	-.03 (.01)	-.06 (.01)
Change in Continental's Scale	-1.78 (.65)	-2.49 (.91)
1994 Performance	-.58 (.03)	-.58 (.05)
Part-Time Service	.57 (.09)	.58 (.12)
Full-Time Service	-.17 (.05)	-.29 (.07)
Weather	-.23 (.06)	-.41 (.07)
Schedule Improvement (Arrivals Improvement)	.21 (.02)	.18 (.03)
New Manager	-1.55 (.44)	
New Manager with Full Outsourcing	1.74 (1.01)	
New Manager with Partial Outsourcing	1.28 (.66)	
Reopened	6.59 (1.33)	9.75 (1.70)

NOTE.—Standard errors are in parentheses. The dependent variable is Improvement. In column 1 the four nonoutsourced airports with the worst 1994 on-time departure performance are omitted, and in column 2 the months following a change in airport management are omitted. The sample size and adjusted R^2 equal 581 and .77 for column 1 and 328 and .81 for column 2.

Higher Initial Performance at Outsourced Airports

On-time departure performance was on average 1.3% better at the 10 outsourced airports than at the 22 nonoutsourced airports. Higher initial performance at the outsourced airports is consistent with Continental's 1994 standing as the worst-performing major airline. However, if it was harder to improve airports that were already performing well, higher initial performance may explain the smaller improvements observed at the outsourced airports. We controlled for this possibility by including a measure of each airport's 1994 on-time performance (1994 Performance) in our analysis. We also reanalyzed the data after rescaling the performance measures using a log odds scale. The findings were unaffected by this change. As further confirmation that our results were not due to differences in initial conditions, we reestimated the model, omitting the four nonoutsourced airports that had the worst 1994 performances. Average on-time performance in 1994 at the remaining 18 nonoutsourced airports was slightly higher (0.14%) than at the 10 outsourced airports. The findings are presented in the first column of table 5 and confirm that the results are robust to this modification. It appears that the smaller improvements associated with outsourcing cannot be explained by higher initial performance at the outsourced airports.

Table 6
Average Supervisor / Employee Ratios

	Partially Outsourced	Nonoutsourced
1994	.107	.107
1995	.096	.081
1996	.102	.082
Sample	6	22

Management Changes

In a recent paper, Gittell (1996) argues that the superior performance in on-time arrival performance enjoyed by Southwest Airlines is partially attributable to relatively high supervisor to employee ratios at that airline.¹⁷ Consistent with this argument, an increase in the supervisor to employee ratio between 1994 and 1996 at the nonoutsourced airports might explain the larger improvement observed at these airports. However, Continental management claimed that the level of supervision at nonoutsourced airports actually fell between 1994 and 1996. As confirmation we gathered annual data on the number of hours worked by Continental supervisors and employees at the nonoutsourced and partially outsourced airports. No data were available at the fully outsourced airports, as employees at these airports are employed by the outsourcing firm. The average supervisor/employee ratios (number of supervisor hours per employee hour) at the partially outsourced and nonoutsourced airports are summarized in table 6. The data reveal a much larger reduction in the supervisor : employee ratio at the nonoutsourced airports than at the partially outsourced airports. It appears that the larger improvement observed at nonoutsourced airports occurred despite a reduction in management supervision.¹⁸

Finally, as additional confirmation that the reassignment of airport managers did not favor the nonoutsourced airports, we reestimated the monthly model using only those months prior to the management changes at each of the 32 airports. After omission of the months following a management change, 328 observations remained. The findings when reestimating the model using these 328 observations are presented in the second column of table 5. The results are very similar to the findings in table 3, suggesting that the changes in airport management cannot fully explain the smaller improvements observed at the outsourced airports.

¹⁷ Gittell's argument is based on the role that supervisors have in facilitating and monitoring cross-functional coordination.

¹⁸ Including the supervision ratios in our multivariate analysis did not yield significant coefficients and had little effect on the other findings.

Summary

We have presented evidence that improvements in on-time performance were larger at nonoutsourced airports than at airports that were outsourced. This evidence is consistent with the claims of Continental employees and senior management who attributed Continental's performance improvement (at least in part) to changes in employee behavior due to the bonus scheme. As support for this interpretation, we identified and excluded several alternative explanations. In particular, we controlled for a variety of airport differences, including factors determining which airports were outsourced. We were also able to present evidence suggesting that reassignment of airport management, changes in the outsourcing status of some airports, changes in the level of employee supervision, and differences in 1994 performance cannot fully explain the result. The different improvements observed at the outsourced and nonoutsourced airports might be explained by other (nonmanagement) changes that systematically affected just one of the groups of airports. For example, the introduction of more efficient equipment or computers at the nonoutsourced airports or labor difficulties at the outsourced airports might offer explanations. However, when asked, senior managers were unable to identify any such systematic changes. They noted that outsourcing partners are carefully monitored and that Continental would respond quickly in the event of adverse changes.

Continental's bonus scheme has been mimicked by at least one of its competitors. TWA also had a history of languishing at the bottom of the DOT's performance ratings, ranking last in on-time arrivals in 1995 and 1996 (a position occupied by Continental in previous years). In June 1996, TWA began paying all employees at director level and below monthly bonuses of up to \$100 for improving its ratings. After the introduction of these incentives TWA improved steadily, and by June 1997 it reached a ranking of first in on-time arrivals—its highest monthly ranking since January 1993. In the 6-month period between May and October 1997 an average of 85.3% of TWA flights arrived on time, compared to just 69.9% over the same period in the previous year.¹⁹ We caution that TWA's employee incentives were accompanied by other operational changes that may have contributed to this improvement.

We conclude that the data are not able to rule out the claim that the bonus scheme was effective, and this explanation remains as a possible (and perhaps likely) explanation for our findings. However, while the outsourced airports provide a natural control group, this was far from a controlled experiment. Other factors may also explain the larger improvements at the nonoutsourced airports. Most worrisome are those

¹⁹ This 15.4% improvement compares with an increase of 2.4% at Continental over the same period.

changes that are managerial in nature and not captured by any of our controls, including the influence of the new senior management team and their efforts to motivate and communicate with the Continental employees. These efforts included introduction of bulletin boards and a quarterly employee magazine, regular voice-mail and video statements from the CEO and increased visibility and accountability of senior managers. Moreover, while the findings suggest that the bonus scheme may have contributed to the improvements in on-time performance, they do not support a conclusion that the bonus scheme alone was sufficient. In a personal conversation discussing our results, the COO succinctly summarized this caution: "There are a handful of critical steps taken to turn around Continental Airlines; the on-time bonus was one of them."

In the next section, we offer an explanation for why the bonus scheme may have been effective. We do so by addressing the threat of free riding and investigating how mutual monitoring may have helped to resolve this threat.

III. Free Riding and Mutual Monitoring

The claimed success of Continental's bonus scheme is surprising. The scheme included approximately 35,000 employees who individually had a negligible influence on overall performance. Moreover, Continental's employees were (are) geographically dispersed, restricting direct interactions and preventing them from observing the efforts of employees at other locations. We might expect these conditions to favor free riding and limit the development of mutual monitoring. In this section, we argue that, despite its size, Continental exploited the benefits of mutual monitoring, thus mitigating the free-riding problem. There are three steps to our argument. First, we argue that structure of airport operations facilitates mutual monitoring between employees. Next we argue that the organization of these airports in a highly interdependent flight schedule supports an equilibrium where each group of employees collectively choose the same level of effort. Finally, we argue that Continental's adoption of the incentive scheme together with other coinciding improvements raised expectations that other airports were improving their on-time performance, enabling movement to the high effort equilibrium.

Free Riding and Mutual Monitoring in Large Firms

An incentive scheme like Continental's is expected to raise efforts only if the additional likelihood of receiving the bonus is sufficient to compensate employees for working harder. Formally, an individual employee deciding whether or not to work hard will do so if and only if $bg > c$, where b describes the value of the bonus, g denotes the incremental likelihood that the bonus is paid when the employee works hard, and c

represents the amount required to compensate an employee for additional effort. Because a single employee has a negligible influence on overall performance, we expect g to be very small. It is this observation that makes the apparent success of Continental's bonus scheme so surprising. In general, we expect g to be a decreasing function of firm size (n) because overall firm performance is less sensitive to the actions of individual employees in large firms.

By paying all employees a bonus based on satisfaction of a common goal, Continental's incentive scheme introduces externalities between the efforts of employees and the welfare of their colleagues. Choice of low effort by any employee not only reduces the probability that he or she will receive the bonus, it also affects the likelihood that other employees will receive the bonus. This creates incentives for employees to monitor their colleagues and encourage them to work harder. Mutual monitoring (sanctioning) can take on two forms. First, there is peer pressure in the form of direct sanctioning of co-workers who do not work hard, together with peer pressure generated by the feeling of shame from not working as hard as colleagues (Kandel and Lazear 1992).²⁰ Second, an employee may sanction a co-worker by reporting low effort to management. Either way, mutual monitoring is an additional penalty imposed on co-workers who work below (or above) the effort norm established by the group.

There are two reasons to doubt that mutual monitoring will support a high effort norm. First, to the extent that it is costly to monitor and sanction co-workers, there is an incentive to free ride on the monitoring and sanctioning efforts of other co-workers. Moreover, just like the first-order free-rider problem, the larger the group the greater the second-order free-rider problem. Hence, the likelihood of effective mutual monitoring is decreasing in both the size of the group and cost of the mutual monitoring. Second, Continental's employees were geographically dispersed, restricting direct interactions and preventing them from observing the efforts of employees at other locations. Both reasons have been used previously to support forceful predictions that mutual monitoring will not occur in large firms.²¹

Mutual Monitoring within Continental's Autonomous Work Groups

Like other large U.S. airlines, Continental is characterized by a large number of autonomous work groups, with one or more work groups at each airport. In this section, we argue that interdependence within and

²⁰ Kandel and Lazear (1992) also consider the impact of guilt, which does not require that co-workers observe an employee's effort.

²¹ See, e.g., Abreu, Pearce, and Stacchetti (1986), Heckathorn (1988), Coleman (1990), and Kandel and Lazear (1992).

between each work group, together with the highly standardized activities that they perform, facilitates mutual monitoring.

The presence of autonomous work groups mitigates the second-order free-rider problem by reducing group size. This makes monitoring more feasible and limits the opportunity to free ride on the monitoring and sanctioning efforts of colleagues. Furthermore, there are reasons to believe that the structure of airport operations dramatically reduces the cost of monitoring. Ramp and gate employees typically work alongside each other performing overt tasks: pushing out or waving in aircraft, loading and unloading baggage, and greeting and boarding passengers. As a result, they are well placed to observe the performance of fellow team members. Moreover, tasks are standardized and rotated, so employees have a common understanding of what behavior distinguishes high and low effort.²² For example, employees unloading baggage from a cart have also spent time inside the hold and can recognize whether delays are due to unusually heavy baggage or neglect by an employee in the hold. Finally, the organization of employees into groups ensures repeated interactions with the same colleagues, revealing habitual (rather than isolated) behavioral patterns.

We might expect that working repeatedly alongside the same colleagues would aggravate sanctioning costs by increasing the cost of damaging social relationships. However, Coleman (1990) argues that in the general context of group norms, sanctioning need not be costly if there is consensus that a high effort “norm” has been violated. Consensus reduces the costs of sanctioning for several reasons. First, it provides social support to the individual that implements the sanction. Second, it may lead to sanctions implemented at the group level (e.g., ostracism from the group’s social interaction). Third, it may increase the likelihood of managerial sanctions against the violator. Finally, the mere knowledge that other group members will discuss a norm violator’s behavior may ensure compliance with the norm.²³

Some of the factors that facilitate monitoring also facilitate consensus. Regular interaction with the same employees provides ample opportunity for information about poor performance to be disseminated. Because the work process is standardized, there is a common and accurate interpretation of poor performance. Finally, a high degree of interdependence between group members causes the costs of norm violation (low effort) to be shared. Interdependence is introduced both by the bonus itself

²² Kandel and Lazear (1992) argue that higher levels of employee homogeneity facilitate mutual monitoring by increasing the accuracy of the monitoring.

²³ Coleman’s (1990) arguments rely in part on Merry (1984) and the role of gossip in supporting social behavior. See also Burt and Knez (1995), who document the impact of gossip and third-party ties on levels of trust among managers in a very large high-technology firm.

(everyone cares about joint output) and the nature of the ramp and gate operations. A flight cannot depart until the entire ramp and gate activities have been performed, so that poor performance by one employee can negate good performance by the rest of the group. For example, maintenance or fueling delays will prevent a flight from leaving on time, even if passengers and baggage are loaded and the plane is otherwise ready to leave. Similarly, if the flight crew is slow to complete their preflight checks, passengers cannot be boarded despite the completion of catering and cleaning activities.

The nature and organization of the gate and ramp activities appear to support our claim that monitoring and sanctioning within Continental's work groups were feasible and may not have been costly. Conditional on a work group agreeing to adopt a high effort norm, consensus that an employee is performing poorly will develop quickly and support sanctioning. Although formal evidence of sanctioning and monitoring is hard to collect, we did learn of several examples that offer further support for our claim.

Anecdotal Evidence

As a part of our initial data-gathering exercise we interviewed the CEO, the COO, three supervisors, five ramp and gate employees, and airport managers at three of Continental's airports (including its largest hub). Four of the ramp and gate employees were interviewed as a focus group. We asked questions concerning the overall structure of operations and observed the process of performing ramp and gate activities. We also posed a series of open-ended questions directed at identifying perceived changes in employee behavior and the reasons for these changes.²⁴ Care was taken to avoid leading questions, especially concerning the bonus scheme. The responses and anecdotes that follow are firsthand accounts of employee, supervisor, and airport manager experiences.

A consistent response from the focus group discussion was the emergence of a team orientation that had been absent prior to 1995. We were concerned that employees were reciting the new corporate "mantra" and so asked for explicit examples of employees behaving "more like a team." We were told that after introduction of the scheme, employees began initiating their own performance reviews whenever airport operations caused a flight to be delayed. These meetings, which were neither attended nor sponsored by managers, focused on identifying sources of delay. Employees also began contacting colleagues who had called in sick, to ask

²⁴ For example, to prompt discussion of general changes in the work environment we asked the following questions: How do you think the work environment has changed? What factors do you think have caused these changes? Which of these factors do you think is most important?

whether any assistance was needed. According to the employees, these calls enabled them to monitor whether the absences were due to valid illnesses.²⁵

We also heard several examples of employees sanctioning each other, including employees being summoned from break rooms by colleagues, employees being chastised for leaving their stations, and gate agents climbing into aircraft holds to identify and help overcome sources of delay. In one example, employees themselves initially attempted to improve a colleague's performance by providing additional training and reallocating his tasks. When this did not resolve the problem, they reported the situation to management, which (after further opportunities to improve) led to that employee's dismissal. An increase in the use of social sanctions was apparently particularly noticeable during the pilots' industrial action in May and June of 1995. The industrial action had a dramatic effect, resulting in rankings of eighth and tenth in on-time performance for May and June (respectively).²⁶ Gate and ramp employee reacted strongly, directly confronting pilots who were causing delays and withdrawing cooperation such as access to break rooms and recommendations for dining and ground transportation.

During the focus group the employees volunteered the bonus scheme as an important factor influencing changes in their own behavior. One employee stated that the bonus plan proved that management was serious, and another employee admitted, in a somewhat embarrassed tone, that \$65 was an attractive incentive. These sentiments appeared to be shared by the other employees.

Inducing Work Groups to Select High Effort

Although there is reason to believe that mutual monitoring occurred within Continental's autonomous work groups, this is not sufficient to explain the claimed success of the scheme (even if the incentive scheme contributed to the development of mutual monitoring). We must also explain why the scheme was able to induce the groups to select higher effort.

As we discussed, an incentive scheme like Continental's will raise efforts only if the additional likelihood of receiving the bonus is sufficient to compensate employees for working harder. When a group acts collectively it considers how joint efforts affect each group member. Whereas an

²⁵ According to the manager of this airport, similar activities had also emerged at other airports.

²⁶ See fig. 1. Recall that these were the only months in 1995 for which the bonus was not paid following introduction of the scheme in February 1995. The airport manager confirmed the employees' response to both the poorly performing colleague and the pilots' industrial action.

individual will work hard only if $bg > c$, the group will collectively prefer high effort when $nbg_n > nc$, where g_n measures the impact of the collective decision on the probability of receiving the bonus. Because this probability is more sensitive to the efforts of a group than to those of an individual ($g_n \gg g$) this condition is weaker than the incentive compatibility condition for an individual, and so the bonus scheme is more likely to induce high effort.²⁷ If Continental's on-time performance is not sensitive to the efforts of the individual work groups, inducing high effort would still require a very large bonus.²⁸ However, there are reasons to believe that Continental's overall on-time performance is sensitive to individual group efforts. When a flight departs late, employees, equipment, and terminal gates are unavailable to service other arrivals and departures. The problem is further compounded when flights carry connecting passengers because departing flights may have to be delayed to allow passengers to make their connections. The capacity to disrupt the entire route system increases the sensitivity of overall performance to individual group effort and reduces the bonus required to induce high effort.²⁹

However, the same interdependencies that increase sensitivity may also hinder change by introducing characteristics similar to a Leontief production function. If a group expects other groups to work hard it can ensure good performance by also working hard. When other groups are expected to perform poorly, output will be low irrespective of that group's decision. In this manner, common expectations of low (or high) effort are both self-fulfilling and reinforcing. This contributed to the problems facing Continental's new management team at the start of 1995. The airline had endured a sustained period of poor performance and (in interviews) many employees recalled being skeptical that the incentive scheme would change behavior of colleagues at other airports. Moreover, after enduring

²⁷ The firm can influence this condition by varying the performance goal and/or the size of the bonus. Moreover, because the probability of attaining the goal depends on the actions of other employees, behavior may depend on how other colleagues are expected to behave. We return to both points in later discussion.

²⁸ This could have led to the well-known budget-balancing problem in which the required bonuses exceeded (in aggregate) the value of joint output (Holmström 1982).

²⁹ Recall from our earlier discussion of outsourcing contracts that several factors prevent the use of airport level incentives. First, scale differences make industry comparisons difficult at the airport level. Second, interdependencies between flights cause on-time performance at one airport to depend on performance at other airports. Finally, rewarding employees for local improvements may cause effort to be misdirected, focusing attention on local improvements to the detriment of the rest of the system. This is well illustrated by the response to an earlier incentive scheme directed at pilots. Prior to 1995 pilots participated in a fuel-saving profit-sharing scheme—the less fuel they used, the more they earned. As a result, pilots responded to on-time departures by flying at slower speeds.

years of bad management, the employees were naturally skeptical of management's ability to take the additional complementary steps required to translate their higher effort levels into higher levels of on-time performance.

Continental's management responded in several ways to overcome this inherent and legitimate skepticism. First, they initially set a modest performance goal, so that small improvements were sufficient to ensure payment of the bonus. Following a sustained period of improvement, the goal was later raised to ensure that the bonus remained sensitive to each group's efforts (see earlier discussion). Second, expectations may have been favorably affected by attribution bias. Introducing the bonus scheme at the same time as the flight schedule improvement led to an immediate improvement in performance, and employees could not easily determine whether the performance improvements were due to the flight schedule changes alone or also reflected improvements in the performance of other employees. The possibility that others were working harder made it optimal to increase effort, which in turn raised the likelihood that others were working harder.³⁰ Third, senior management undertook several initiatives to improve communication with employees.³¹ Under these initiatives, senior management described their strategy for improving performance and responded to employee questions and complaints. Together with the flight schedule improvement, it is reasonable to conjecture that these efforts increased the new management's credibility and created common expectations (across airports) that performance would improve.

Mutual monitoring does not explain why the bonus was extended to employees who do not affect on-time performance, including, for example, telephone reservation personnel. When we asked why the scheme was offered to all employees and not just employees that affect on-time performance, the CEO responded (quite forcefully) that it was important that all employees understand that everyone's effort and commitment to improvement is critical. This may be interpreted as a strategy for reinforcing senior management's commitment to improving on-time performance by changing employee behavior.

Summary

We have argued that the claimed success of Continental's incentive scheme was partly due to employees monitoring each other. Although there are obstacles to monitoring in large firms, Continental was appar-

³⁰ Under this interpretation the response of the employees to Continental's bonus scheme is an example of a subjective equilibrium in a repeated game (see Kalai and Lehrer 1993).

³¹ We briefly reviewed these initiatives at the end of the previous section. For additional details, see Brenneman (1998).

ently able to overcome these obstacles due to the division of employees into autonomous work groups and the nature of its ramp and gate activities. Collusion within the work groups effectively reduced the agency problem from 35,000 individual workers to a single agent at smaller airports and a handful of groups at larger airports. We claimed that interdependencies between the performance of the different groups both helped and hindered the incentive scheme. The interdependencies helped overcome the lack of group-level measures by making overall firm performance more sensitive to individual group efforts. However, by increasing the dependence of groups on each other, they also reinforced low effort. Continental appears to have responded to this hindrance by introducing the bonus scheme together with other changes and varying the performance goal. Although the initial goal was attainable even if some groups did not improve, it was later raised so that it continued to be sensitive to individual group decisions.

Other explanations that may have contributed to the claimed success of Continental's incentive scheme include "efficiency wage" and "implicit incentives" arguments. The "efficiency wage" argument posits that raising net compensation increases the opportunity cost of being dismissed for shirking (Akerlof and Yellen 1986). Although this cannot explain evidence that monitoring and sanctioning activities were heightened following introduction of the bonus program, it seems clear that employees were more satisfied in 1996 than in 1994, as evidenced by the reported reduction in employee turnover. Additional satisfaction represented a premium that may have prompted increased concern for retaining employment.

Theories of "implicit incentives" recognize that not all compensation is linked to explicit compensation. Promotions and social recognition may provide strong incentives, yet they rarely depend on explicit criteria or deterministic decisions. It is possible that Continental's incentive scheme helped to reveal the criteria used to apportion implicit rewards, so that increased effort was not motivated solely by the explicit monthly bonus. However, this argument faces two limitations. First, the importance of on-time performance was not surprising. Employees almost certainly already recognized that actions taken to improve on-time performance would influence the allocation of implicit rewards. Second, baggage, customer complaint, and other performance measures also improved, suggesting that employee effort was increased rather than simply reallocated.

A more behavioral view of employee behavior might argue that we place too much emphasis on the role of mutual monitoring. Employees of Continental had endured years of poor working conditions and low performance, and there are presumably nonmonetary benefits in working for a successful organization. Continental employees could have been induced to work harder because they perceived that they were now part of a successful organization. Under this more behavioral view, the firm-

wide incentive scheme is one of several measures that helped raise performance expectations.

IV. Conclusions

We isolated the impact of the firm-wide bonus scheme introduced by Continental Airlines by comparing the change in performance at outsourced and nonoutsourced airports. We found evidence that performance improvements were larger at nonoutsourced airports, which is consistent with the bonus scheme raising the efforts of those employees eligible to receive it. The result is robust, surviving several tests designed to rule out alternative explanations. However, the apparent success of the program is surprising: we might have expected that the incentives created by the scheme were too weak to prevent free riding. Our explanation for this surprising result focuses on how mutual monitoring may mitigate free riding by forcing employees to internalize the impact of their actions on each other. In support of this explanation, we argued that the nature and organization of Continental's operations were particularly conducive to the development of mutual monitoring between employees.

References

- Abreu, Dilip; Pearce, David; and Stacchetti, Ennio. "Optimal Cartel Equilibria with Imperfect Monitoring." *Journal of Economic Theory* 39 (1986): 251–69.
- Akerlof, George, and Yellen, Janet. *Efficiency Wage Models of the Labor Market*. New York: Cambridge University Press, 1986.
- Alchian, Arman A., and Demsetz, Harold. "Production, Information Costs, and Economic Organization." *American Economic Review* 62 (December 1972): 777–95.
- Blasi, Joseph; Conte, Michael; and Kruse, Douglas. "Employee Stock Ownership and Corporate Performance among Public Companies." *Industrial and Labor Relations Review* 50 (October 1996): 60–79.
- Brenneman, Gregory. "Right Away and All at Once: How We Saved Continental." *Harvard Business School* 76 (September–October 1998): 162–79.
- Burt, Ron, and Knez, Marc. "Kinds of Third-Party Effects on Trust." *Rationality and Society* 7 (1995): 255–92.
- Coleman, James S. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press, 1990.
- Conte, Michael A., and Kruse, Douglas. "ESOPS and Profit-Sharing Plans: Do They Link Employee Pay to Company Performance?" *Financial Management* 20 (Winter 1991): 91–100.
- Cooke, William H. "Employee Participation Programs, Group-Based Incentives, and Company Performance: A Union-Nonunion

- Comparison." *Industrial and Labor Relations Review* 47 (July 1994): 594–609.
- Gittell, Jody H. "Coordination, Control and Performance of Interdependent Work Processes." Working Paper no. 96-042. Cambridge, MA: Harvard Business School, 1996.
- Heckathorn, Douglas D. "Collective Sanctions and the Creation of Prisoner's Dilemma Norms." *American Journal of Sociology* 94 (November 1988): 535–62.
- Holmström, Bengt. "Moral Hazard in Teams." *Bell Journal of Economics* 13 (Autumn 1982): 324–40.
- Kalai, Ehud, and Lehrer, Ehud. "Subjective Equilibrium in Repeated Games." *Econometrica* 61 (1993): 1231–40.
- Kandel, Eugene, and Lazear, Edward P. "Peer Pressure in Partnerships." *Journal of Political Economy* 100 (1992): 801–17.
- Kruse, Douglas L. "Does Profit Sharing Affect Productivity?" Working paper. Piscataway, NJ: Rutgers University, June 1993.
- Merry, S. E. "Rethinking Gossip and Scandal." In *Toward a General Theory of Social Control*, edited by D. Black, pp. 271–302. New York: Academic Press, 1984.
- Welbourne, Theresa; Balkin, David; and Gomez-Mejia, Luis. "Gainsharing and Mutual Monitoring." Center for Advanced Human Resource Studies Working Paper no. 93-12. Ithaca, NY: Cornell University, 1993.