The Contingent Effect of Management Practices

Steven Blader
New York University

Claudine Gartenberg
New York University

Andrea Prat
Columbia University

Initial draft: October 15, 2014*

** Preliminary and incomplete. Please do not distribute. ***

Abstract

This paper investigates how the success of a management practice depends on nature of the relational contract between the firm and its employees. A large US logistics company is in the process of fitting its trucks with an electronic on-board recorder (EOBR), which provide drivers with information on their driving performance. In this setting, a natural question is whether the optimal managerial practice consists of: (1) Letting each driver know his or her individual performance only; or (2) Also providing drivers with information about their ranking with respect to other drivers. The company is also in the first phase of a multi-year "lean-management journey". This phase focuses exclusively on changing employee values, mainly toward a greater emphasis on teamwork and empowerment. The main result of our randomized experiment is that (2) leads to better performance than (1) in a particular site if and only if the site has not yet received the values intervention and worse performance if it has. The result is consistent with the presence of a conflict between competition-based managerial practices and a cooperation-based relational contract; it also highlights the role of relational contracts in determining the success of management practice.

*We thank...

1 Introduction

Economists have increasingly focused on management practices as an important explanation for the large observed variation in productivity among firms. In an early paper on the subject, Ichniowski, Shaw and Prennushi (1997) find that improved HR practices are associated with up to 10% greater uptime between essentially identical steel finishing lines. Subsequent research has generally corroborated this pattern. For instance, recent U.S. census data of manufacturing plants analyzed by Bloom et al (2014) yields a strong relationship between management practices and an array of performance measures. Experimental field studies have also investigated similar effects, primarily in developing markets (e.g., Drexler, Fisher and Schoar 2011; Bloom et al 2013; Karlan, Knight and Udry 2014).

A straightforward question follows from these findings: if there are such clear returns to management practices, why do large differences persist across firms, even within narrowly defined industries? One proposed answer to this question is that poor institutions make practice adoption costly and also protect inefficient firms from competition (Bloom et al 2013). While providing a convincing argument for firms in developing countries, this explanation does not explain the large observed dispersion in wealthier settings or the substantial within-country variance, where firms presumably face similar institutional contexts.

A second proposed answer, explored by Gibbons and Henderson (2013), is based on relational contracts. A relational contract is a non-legally binding understanding between a firm and its employees, which typically describes how employees should behave and how the firm will in turn reward the expected behavior (MacLeod-Malcomson 1989, Baker et al 2002; Levin 2003). The promises of the two parties are enforced through repeated interaction. Gibbons and Henderson (2013) note that the relational contract of a particular company is difficult to change. Hence, similar firms may be governed by very different relational contracts. However, the ability of a firm to introduce a new management practice is likely to depend on the relational contract that is in place, which can explain the persistent heterogeneity of management practices even within the same country and industry. This explanation was alluded to indirectly by Ichniowski, Shaw and

Prennushi (1997) in their discussion of practice adoption among steel finishing lines. They speculate that differences between lines could be partly explained by different levels of trust between labor and management. In support of this explanation, they note that the best practices were observed either in newly constructed lines or in lines that had been shut down and re-opened under new ownership.

This hypothesis is difficult to test directly for two reasons. First, almost by definition, relational contracts are difficult to observe because they contain prescriptions that cannot be expressed in a legally binding way. Second, even if they were observable, given that each relational contract is associated with a company, it would be difficult to disentangle the effect of the relational contract from other firm-specific unobservable factors.

The goal of this paper is to make some progress on this issue by studying a company that is arguably transitioning from one relational contract to another. This transition is implemented on a site-by-site basis across the company and is a lengthy process. Since the company was mid-way through the transition at the time of our experiment, some locations within the firm had a version of the new relational contract, while others did not. We study how workers in these different locations react to the introduction of the same management practice.

Specifically, we run a field experiment within a trucking company with a large number of sites that all perform a similar function throughout the United States. Our company recently introduced electronic on-board recorder (EOBR) technology that measures the performance of drivers against a route-specific benchmark. A natural question for the firm is whether drivers should only be made aware of their individual performance (no disclosure) or whether they should also learn about their colleagues' performance (relative performance disclosure).

The company is also implementing a far-reaching change to its management culture. They are engaged in a multi-year program to roll out Lean Management. Lean Management ("lean") is widespread management philosophy, inspired by the Toyota Production System, centered on teamwork and worker empowerment. Gibbons and Henderson (2013) characterize the Toyota system as a relational contract. At our company, a successful lean implementation requires profound changes in employee behavior across all levels of the organization, given the prior management

culture that had been at place at this firm for many years. Accordingly, the company has committed substantial resources to the implementation and a ten-year schedule for the rollout.

This rollout is composed of five phases and, at the time of the experiment, the company was midway through the first phase, which involves no change in the operations of the company or the incentives of the employees. It mainly consists of instilling Lean principles in employees, particularly how "continuous improvement" (the organizing idea behind Lean Management) occurs through teamwork, collective effort and the empowerment of front-line workers. In a relational contract framework, this first phase can be seen as a costly effort on the part of the company to announce a change in the relational contract. For the purpose of this paper, we will refer to this first phase as the "lean intervention" or "lean," even though Lean Management, in its fullest sense, involves many other changes to both formal and informal operating practices (which had fortunately not yet been initiated at our company). At the time of the experiment some sites had received the lean intervention while others were untouched.

Our main finding is that the effect of whether drivers' performance was publicly posted strongly depended on whether the sites had received the lean intervention. Drivers assigned to untouched sites responded on average positively to the performance postings, improving their fuel efficiency by 3.75% and reducing their idling time and wasted fuel respectively by 1.3% and 1.6% relative to the control group. In contrast, drivers at lean sites responded negatively to the individual performance rankings. We record a substantial drop in performance for these drivers, in the form of an 13.4% reduction in fuel efficiency and an increase of 3.5% in idling time and 5.4% in wasted fuel, relative to the control group.

These results are consistent with relational contracts affecting practice adoption. This explanation is supported by several streams of research in social psychology that predicts differences in individuals' behavior driven by whether they possess collectivistic or individualistic orientations towards their groups (Brewer, 1991; Brewer & Gardner, 1996; Markus & Kitiyama, 1991; Tajfel & Turner, 1978; Triandis, Bontempo, & Villareal, 1988). In this work, a collectivist orientation reflects a model of oneself as bonded with the group as an overall entity. This cognitive model, in turn, prompts the sharing of common objectives. In contrast, an individualistic orientation, consistent

with common assumptions within economics, reflects a model of oneself as distinct from the other group members and possessing individual goals. This research has found that these orientations fundamentally shape cognition and behavior, with one consequence that adverse reactions are likely when management practices are inconsistent with workers' dominant orientation.

Based on this research, we interpret our results as follows: within the untouched sites, an individualistic orientation prevails, reflecting the firm's original management culture. Public performance posting is therefore perceived positively since it provides drivers an opportunity to distinguish themselves from each other. It motivates increased effort via social incentives such as public individual recognition (or shame from inferior performance) .

In contrast, within lean sites, a collectivistic orientation dominates as a result of the emphasis on teamwork, collaboration, and empowerment. Therefore, public performance rankings are interpreted as pitting "team member against team member", and directly violating the collectivistic contract facilitated by the company's management. Organizational psychologists have found that violations of these informal "psychological" contracts are perceived negatively by workers, undermining their motivation and their commitment to the organization (Rousseau, 1995). In addition, combining individualistic and collectivistic management practices has been shown to convey mixed-messages about what is value and to sow confusion among employees (Wageman, 1995). The results of our study reflect these two patterns. While we cannot confirm this interpretation directly, we do run several tests whose results are consistent with the notion that the difference between individualistic and collectivistic orientations underlies the different responses of lean and non-lean sites.

The biggest challenge in our experimental design is the assignment of the lean intervention across sites. Ideally, both the performance postings and the lean intervention would be randomly assigned to sites. However, the constraints of our study required us to use a quasi-random assignment for the lean intervention: the posting treatment had a specific three month window to be rolled out across all the sites, while the initial phase of lean implementation was scheduled to occur over a five year period across all sites, with a minimum of three to six months to complete at any given site. Given this timing mismatch, we instead stratified our random assignment of performance postings by

whether a site had received the lean intervention at least three months prior to the commencement of the study. We further applied statistical matching between sites with and without lean intervention. Importantly, our results generally strengthened or remained stable after the matching.

It is plausible that this empirical challenge is unavoidable given our research question. If relational contracts could be costlessly and immediately implemented, then we might expect much more convergence of both relational contracts and management practices among firms. Although we cannot show this in our study, we speculate that it is complementarities between these costly, slow-to-implement relational contracts and simpler, formal management practices that contribute to the high dispersion of practices across firms.

This study has several implications. The most direct one relates to the particular management practice under study, the publication of relative performance rankings. There a number of experiments on the effect of relative feedback interventions (see for instance Bandiera et al 2012, Barankay 2012, Blanes i Vidal and Nossol 2011). The results are mixed, with some studies showing an improvement in performance and others showing a decline. Within this literature, our contribution is to identify that the relative benefit of not disclosing ranking information can depend on the predominant relational contract between management and workers.

A second conclusion of the paper is that the mechanism hypothesized by Gibbons and Henderson (2013) appears to be consistent with our evidence. Different relational contracts can lead to differences in how workers respond to the introduction of new management practices. A practice that is beneficial under one relational contract can be detrimental under another one. This in turn has two implications: it may explain why seemingly identical firms adopt different practices and why practices that seem very beneficial for certain firms are not adopted by other ones.

Third, the study raises a methodological question about the generalizability of findings from field experiments, a rapidly expanding method within economics (see Bandiera, Barankay and Rasul 2011 for a survey). Specifically, the result of a field experiment conducted in one organization may not be a good predictor of the results of the same experiment in other similar organizations or even in the same organization at a later stage. Using our experiment as an example, had we run our study in 2010, prior to the commencement of the lean initiative, and omitted any consideration of the

relational contract of each site, we would have concluded that performance postings have a positive impact on productivity. Alternately, had we run our experiment in 2018, after the completion of the initial lean phase, we would have concluded the opposite. In this sense, our study is an invitation to provide more consideration of conditions under which the findings of field experiments are generalizable.

2 Literature Review (TBD)

3 The Research Setting

3.1 Why the Trucking Industry?

The US trucking industry has several features that make it well suited for research on relational contracts and management practices. Intense competition and well developed information markets (in the form of trade organizations, conferences and consultants) lead firms to rapid adoption of productivity-enhancing technology. A subset of this technology provides managers with extensive data and monitoring capabilities, enabling them to implement a broad range of previously infeasible operating practices. Among the firms that adopt this new technology, there is also broad scope to implement new practices. In fact, managers are effectively required to do so, given that the technology is only useful insofar as it is effectively integrated into the daily operations of the company. In this sense, we can view these new technologies as an outside shock to management practices across the industry.

One of these technologies is of particular importance to our research design. Electronic on-board recorder systems (EOBR) (cite Baker Hubbard here?) record and transmit detailed driving behavior to a centralized database accessible to managers. This database can be used to evaluate, discipline and reward drivers in near-real time. EOBR systems also include terminals installed in truck cabs that display driver performance information and emit audible real-time alarms when driving behavior is determined to be inefficient.

A second feature of this setting is that the new technology and associated practices can be

viewed by drivers as highly intrusive (in fact, at the time of our writing this study, a new technology was announced to install cameras that measure the height of drivers' eyelids to gauge their fatigue). If implemented improperly, firms run the risk of alienating their workforce, which can result in reduced productivity, sabotage of the changes and unionization. From our discussions with company management, this was their primary concern as they decided when and how to roll out new technology. In this sense, the new operating practices can be viewed as complementary to the relational contracts between managers and drivers at these companies.

Related to this point is that the industry has a long history of driver independence: companies have traditionally allowed a high degree of independence to their "last American cowboys," in exchange for long hours and monotonous work.[1] In this sense, the wave of new technology represents a threat to this culture and companies are faced with how to handle this transition smoothly.

3.2 The Company

The company at which we conducted this study operates in the less-than-truckload segment of the industry, transporting shipments that are smaller than full truckload freight and larger than individual parcels. At the time of our study, the company employed approximately 10,000 drivers, all of whom were non-union hourly employees, across approximately 300 sites in the US and Canada. Important for our experimental design, most drivers operate local routes and there is little communication between sites. Shipments are picked up and delivered during regular business hours via local routes of less than 300 miles covered by drivers that can remain with the same customers over months or years. Intercity shipments are transported between sites via by a minority of "line-haul" drivers, typically on an overnight shift. Because of the difference in shift schedules and the small proportion of line-haul drivers, the threat of cross-contamination between sites during our study is limited. This features enables us to establish a credible control group in the experiment.

The company was engaged in two major initiatives at the time of our study that we incorporated into our research design. First, beginning in the second half of 2014 and continuing over a four month

[1] [Trucker culture has been defined by] "the sense of fierce independence, counter-cultural defiance, and unapologetic masculinity...truckers very much valued (and continue to value) not being confined within the four walls of a factory or an office" http://freakonomics.com/2009/02/27/ask-an-economist/)

period, EOBR was rolled out to all trucks. This rollout represented the first time the company managers had information on individual driver's efficiency and were sensitive to how their use of this data would be accepted by the workforce. Accordingly, they were open to experimentation on certain practices as a means to decide how to integrate the technology into their daily operations.

Second, beginning in 2011 and continuing during our study period, the company was engaged in a decade-long program to change their business culture and operations to conform to Japanese manufacturing practices. At the time of our study, the company had initiated the first of a five phase transition to this "lean" operations model, with a plan to complete the first phase across the remaining sites by the end 2015. The focus of the first phase is primarily on transitioning the culture of the site from an individualistic and hierarchical culture to a team-based one.

The timing of this initiative had two advantages for our study. First, the first phase of this transition was primarily focused on changing the prevailing relational contract and involved no changes to formal driver-related workplace practices that could otherwise affect our performance measures. The second through fifth phases did focus on the formal tools side of lean manufacturing, but importantly, none of these phases had been initiated at the time of our study.

Second, 35% of the sites had begun this cultural initiative at least three months prior to the beginning of the study, enabling a meaningful comparison between sites that had undergone the initiative and sites without any culture shift.

Third, after the initial pilot phase, the rollout of the lean initiative was generally decided by the simplest travel schedule of the various managers in charge of training. We consider the rollout, therefore, to be quasi-random for our purposes, in the sense that the rollout schedule is unrelated to the anticipated success of the initiative or other factors that may influence the acceptance of the rank postings.

4 The Experiment

The experiment occurred between August 2013 and July 2014 as EOBR was rolled out throughout the company. We adopted a three by two research design where we assigned three ranking conditions

randomly across lean and non-lean sites.

4.1 The Ranking Intervention

We designed two ranking treatments in addition to the control group: one in which the driver names were posted next to performance information and one in which the employee IDs were used. In this latter treatment, a driver can identify his own standing and view the distribution within the site, but does not know any other individual's rank, nor do others know his rank. We make use of this latter condition in later sections when we provide evidence in support of the psychological mechanism that we propose drives our main result. Because of the substantial number of sites and the lack of pre-existing outcome data to to perform power analyses, we placed equal numbers of sites to each of the three conditions (control, named and IDed postings).

The rankings were posted on a weekly basis, beginning six weeks from the EOBR rollout date for a given site. This timing allowed us to obtain thirty days of pre-measures (we discarded the first two weeks while the systems were calibrated to the trucks). The pre-measures, combined with the control group and lean stratification, enable a triple-differences research design that we describe in Section 5.2.

The postings contain the employee identifier (either driver names or employee IDs, depending on the treatment assignment) and four performance metrics recorded by the EOBR system. These metrics are gap score, shift score, excess idle time and total fuel lost. We discuss each of these in more depth below in Section 4.4. See Appendix Figure A1 for a sample of the posting. The ranks are determined by the gap score

4.2 The Lean Initiative

The company divided their lean launch initiative into five phases, only the first of which had been initiated, and only across approximately 35% of the sites. This initial phase was designed primarily to instill the lean manufacturing philosophy in the workers of each site and, in particular, the role of teamwork and "servant leaders" in enabling continuous manufacturing. Very few formal processes were introduced during this initial phase. Instead, workers at each site went through training on

the ideals and principles of lean manufacturing and organized a "lean team" comprised of front-line workers (drivers, dock workers and others in non-supervisory or management roles). This team underwent additional training and was responsible for modelling "lean values" at the site. This meant, among other activities, having the drivers run their own meetings and work as a team to reorganize the community area and dock as they chose. Appendix Figure A2 shows the criteria by which sites are evaluated after completing this initial phase, which show an emphasis on "soft" changes, such as the nature of the employee-manager relationship and the nature of teamwork at the site. Appendix Figure A3 shows excepts from interviews with drivers and supervisors on the impact of the lean initiative at their sites. These excerpts show that, while workers noticed very few formal changes, they did have a strong sense that the degree of teamwork and management style had both improved as a result.

For an ideal experiment, lean status would be randomly assigned, similar to the ranking conditions. However, this assignment was impractical given the constraints of the study timing and the long timeframe and high management intensity required. The study was constrained to coincide with the EOBR rollout during the last four months of 2014 since rank postings were bundled into the larger EOBR rollout initiative by the company. In contrast, altering the relational contract from an individualistic orientation to a team focus requires at minimum three to six months per site and could not be implemented on a large enough scale to be integrated with the study. Instead, we stratified our ranking assignments by lean rollout status as of the start date of the study and applied statistical matching techniques to correct for differences between lean and non-lean sites.

4.3 Sites Included in Field Experiment

The firm has a total of approximately 275 sites. Of these sites, we discarded 25 from the sample that launched the lean initiative in 2011, the launch year of the program. In discussions with the company senior management on the need for quasi-random assignment of lean, they mentioned that the earliest sites were selected for either exceptional leadership or workforce or other conditions directly related to the anticipated success of the rollout within those sites. While the company selected less than ten sites for these initial pilots, we conservatively discarded all sites launched

in that first year. We further excluded 36 sites that were scheduled to launch lean during the timeframe of the study, as these locations could not be reasonably be classified as either lean or non-lean. Lastly, we discarded 72 of the sites with rankings launched in the first half of the study. Until the mid-project checkpoint, no formal verification process of the rank postings had been instituted and upon further investigation, we learned that there was minimal compliance up to that point. After our discussion, the company instituted a formal process to verify that ranks were posted as required by the experiment guidelines, include weekly photographs of the postings, conference calls, and a shared spreadsheet tracking system.

After these corrections, the experimental sample included 142 sites, 47 in the control group, 50 in the named postings group ("Treatment 1") and 45 in the IDed postings group ("Treatment 2"). To construct our driver-day sample from the EOBR system, we then further removed line-haul routes (defined as routes above 300 miles) and routes with EOBR data that was clearly unreasonable (less than 15 mile routes or MPG<1 or >15). This left us with a sample of 330,689 driver-days.

Because the company did not have the managerial bandwidth to reinforce the ranking messaging on an on-going basis, we expected to see some diminishment of any effects of the postings over time. For our main multivariate analyses, therefore, we restricted the windows of the experiment from the thirty days prior to the postings to the thirty days after. We also removed the five days immediately surrounding the scheduled posting dates, since many of the site managers chose a different day of the work week to post the rankings to coincide with group meetings, rather than on the date specified by the experiment. We were left with a sample of 93,913 driver-days within these narrowed windows that we use in our primary multivariate analyses, although in the appendix we repeat the analyses with the long windows and show that most of the results are replicated, although with somewhat larger standard errors.

Appendix Table A1 contains a summary of the sample construction process.

4.4 Outcome Variables

We focus on four outcome variables for this study that capture different aspects of efficient driving performance. Gap score, the primary measure used in the driver rankings, calculates the difference

between the average actual and "potential" miles per gallon expended on a given route. The potential miles per gallon is calculated by the EOBR system based on what it considers to be optimal shifting and speed patterns, given weather conditions and route characteristics. Gap score is represented in percentage terms such that, for example, if actual and potential mpgs for a given route are 6.5 and 7.0, respectively, the gap score would be 7.7 ((7.0-6.5)/6.5*100). A higher gap score, therefore, represents worse (less efficient) driver performance.

Shift score is the percent of shifting events performed on the route that remains within designated RPM limits for the engine. For example, if a driver shifts five hundred times on a given route, his shift score will be 90 if he revs the engine above a designated threshold during fifty of those shift events. In contrast to gap score (and the two other measures described below), a higher shift score represents better driver performance

Excess idle time is a measure of the minutes an engine idles beyond a designated time period, thereby wasting fuel. This metric particularly captures instances in which the driver allows the engine to idle while making a delivery, counter to company policy.

Lastly, fuel lost is an aggregate measure of all the fuel wasted from idling, inefficient shifting, speeding and gearing. As with gap score, a higher value for excess idle or total fuel lost represents worse performance.

Because all four measures are included in the weekly rank postings, we investigate each of them as outcome variables in our analysis, although we focus relatively more attention on gap score since this is the measure by which ranks are determined.

Table 1 reports summary statistics for the sites in the field experiment. Note that the sample of 142 sites used in the experiment is representative of the full set of 275 sites within the firm, based on observable site characteristics and pre-rank posting driver performance. Within the sample, the Control and Treatment 1 (named postings) groups are statistically indistinguishable, while the Treatment 2 (IDed postings) group appears somewhat different from the other two groups, particularly in the distribution of lean sites and pre-posting driver performance. We control explicitly for these differences in our primary analysis and also perform additional analyses using a matched sample. However, even with these corrections, we interpret results concerning the Treatment 2

group with more caution than those of the Treatment 1 group.

5 Impact of Rankings and Collective Values on Driver Performance

5.1 Preliminary evidence

Figure 1 depicts the combined response of both lean and non-lean sites to the driver postings, as measured by the each of the four outcome variables. The x-axis is normalized such that week 0 represents the week that rankings were posted at each site, regardless of the calendar date of each posting. The y-axis measures the outcome, with higher values generally signifying worse performance (with the exception of shift score where higher values signify better shifting behavior). We aggregate the data in two complementary ways. In Figure 1a, we plot the median values of each outcome and weight each site equally, regardless of the number of drivers per location. In Figure 1b, we plot the average values of each outcome variable across all drivers, effectively weighting each driver equally regardless of site assignment.

Four results can be drawn from this figure. First, within the control groups, the performance outcomes either deteriorate or remain relative constant over time after the EOBR system is launched (at week -6). For example, in Figure 1a, gap scores for the control group range between 1.2 and 1.35% for weeks -6 to 0, and increased to 1.3 to 1.49% for weeks 1 to 18. A similar pattern is apparent for both shift score and fuel lost. According to our field interviews, this trend may be attributable in part to the reluctance of management to assign any formal incentives to driving performance, which may in turn have led to reduced focus on the EOBR system over time.

Second, the pre-posting trends appear similar between Treatment Group 1 (named postings) and the control group, consistent with successful random assignment of sites between groups. Third, after rank posting at week 0, there is a clear divergence in Figure 1a between the Treatment Group 1 and the control group, with the exception of the fuel lost metric. This pattern is not apparent in Figure 1b, which plots the outcomes as averages across all drivers. The difference between Figures

1a and b can be explained by the greater number but smaller sizes of the non-lean sites relative to the lean sites. Figure 1a, therefore, places more weight on the response of non-lean sites to the treatments and show that for these sites, the named posting treatment had an overall positive effect on driver performance. We explore this result further in Figure 2.

The final result apparent from Figure 1 is that Treatment Group 2 (IDed postings) appears to be different from the two other groups in pre-measures of gap score. Although these differences are not statistically significant, we would have preferred to see closer pre-trends between the two groups. Because of these potential differences, we focus more on the named posting group for our analyses and we draw conclusions from the IDed posting group with more caution and only in conjuction with other tests.

<< Insert Figure 1 about here >>

Figures 2a and b depict the performance response to the ranking postings by lean and non-lean sites and show graphically one of the main results of the experiment. For visual clarity, we replace the absolute levels of the performance metrics as shown in Figure 1 with the delta between the treatment and control groups. As such, the 0-line on the Y-axis represents no difference between the two groups, a positive difference (shown by a line above the 0-line) represents worse performance for the treatment group and, conversely, a negative difference (shown by a line below the 0-line) represents better performance (note that, for shift score the opposite is true).

Three results are apparent from this figure. First, each of these four plots shows a clear separation in response between lean and non-lean sites to named postings, with the non-lean sites in the named rank group showing relatively better performance than the control group (above the 0-line) and the lean sites showing relatively worse performance (below the 0-line).

Second, the difference between the control and Treatment Group 1 (named posting) appears somewhat persistent over the timeframe of the experiment, although near the final weeks of each plot, the performance measures converge somewhat. Regarding this convergence, two points are important to note: there is more noise in these final weeks than in the earlier weeks, since only the sites with the earliest ranking rollouts had data that extended this far at the time of the analysis Second, the company did not have the managerial capacity to reinforce the ranking

postings consistently throughout the study period, so we expected some habituation and reversion . As a result, it is not possible to infer whether apparent convergence is an artifact of the data and experiment or is a more general finding.

<< Insert Figure 2 about here >>

Overall, this graph shows preliminary evidence for one of the main results of the experiment: that driver response to social incentives is contingent on the nature of the relational contract of the site to which he is assigned.

5.2 Intent to Treat Estimates

We now turn to multivariate analyses to verify the preliminary results discussed above. We estimate the differential impact of the rankings on lean and non-lean sites using the following triple-differences equation:

$$PERF_{it} = \alpha TREAT\_GROUP_i * LEAN_i * POST_{it} + D'_{it}\beta + e_{it} \qquad (1)$$

where i represents a given driver and t is the calendar date. $PERF$ is one of the four performance outcomes, $TREAT\_GROUP$ is a vector of two indicator variables, one for each of two possible treatments assignments (named or IDed rank postings), and $LEAN$ is an indicator variable that is equal to one if the sites have launched lean at least three months before the beginning of the experiment. $POST$ is equal to one after the assigned date of the ranking rollout for the two treatment groups or, equivalently, six weeks after the EOBR rollout for the control group. All pair-wise interactions and individual variables associated with the triple-differences term are also included in the model and α represents the vector of coefficient estimates for all the associated terms. We are primarily interested in the coefficient on the triple interaction itself, which estimates the difference in response to the rank postings of lean and non-lean sites. D_{it} is a vector of control variables that includes the total number of tractors at the site to measure the size of a site, day of week indicators to absorb weekly patterns, lean manager fixed effects, regional fixed effects, and the distance and potential MPG of the route.

We also perform several variations of this analysis. We add in driver and date fixed effects to control both for driver traits and seasonality. We also create a subsample of the data that matches

lean and non-lean sites to account for the non-random assignment of lean in the experiment. Lastly, we do an instrumental variables analysis where we instrument X with Y [fill in] to account for incomplete compliance. We discuss each of these analyses in turn below.

5.2.1 Combined Effect Across All Sites

We begin by estimating the simple intent-to-treat model without differentiating between lean and non-lean sites. Table 2 shows the results of all four performance outcomes, both without and with controls (odd and even columns, respectively). Consistent with Figure 1b, the ranking intervention appears to have no effect. Without accounting for the underlying values at the site, therefore, we might inaccurately conclude that rank postings have no effect on worker performance.

<< Insert Table 2 about here >>

5.2.2 Effect by Lean and Non-Lean Sites

We next estimate the differential impact between lean and non-lean sites. Table 3 shows the results of the intent-to-treat analysis modeled in Equation (1) and is consistent with the plots in Figure 2. In Columns (1) and (2), we see a large, positive difference in the response of lean sites to the named posting treatment. Using the estimates in Column (2), we observe a 13.4% greater average gap score within lean sites with named postings, relative to control, and a 4.23% lower average gap score within non-lean sites with named postings.

No similar effect was estimated for the second treatment group that posted the IDed ranks. Consistent with successful randomization, we estimate no statistical difference between treatment groups and control groups (coefficients on Treatment Group 1 and 2 and Lean*Treatment Group 1 and Lean*Treatment Group 2), nor any difference between the control groups in the post-rank posting response (coefficient on Post*Lean) We do, however, see some evidence of overall performance deterioration for the control group (coefficient on Post), possibly reflecting the earlier observation by company management that the lack of explicit performance incentives may have led to decreased attention paid to the EOBR system over time. Similar results are evident in Columns (3)-(8) for the other three performance outcomes.

Interestingly, the deterioration in the driving performance within lean sites in Treatment Group 1 (named posting) is consistently more significant and larger than the improvement in the non-lean sites. For example, the reduction is shift score of 0.845 points (10% of a standard deviation) from the Column (3) estimate is larger than the 0.0968 point improvement (1% of a standard deviation) estimated for the non-lean sites.

Also notable here and in subsequent analyses is that the inclusion of control variables does not change the estimates substantially, consistent with a successfully randomized experiment.

<< Insert Table 3 about here >>

Table 4 repeats the analysis including date and driver fixed effects and the results remain largely unchanged. Column (1) estimates a 10.7% lower gap score within Treatment Group 1 lean sites relative to control. As in Table 3, we observe no underlying pattern for Treatment Group 2, the IDed rank group. The similarity in results between Tables 3 and 4 supports the conclusion that the effect of ranking is due to changes in driver behavior and not due to compositional differences between sites.

<< Insert Table 4 about here >>

5.3 Matched analysis

One challenge of the study is that lean status is not randomly assigned. A partial mitigation to this concern is that the company's management indicated to us that - after the 2011 pilot period (not included our sample) - the choice to launch lean at a given location was driven by geographic expediency and schedule optimization for the managers involved in the initiative. However, as is evident from statistics in Table 5, this rollout strategy still resulted in underlying differences between lean and non-lean sites, with the biggest difference between the two group being the average site size. For our experiment, this size difference poses a challenge in interpreting our main result if large sites are both lean and more likely to resist rank postings for unrelated reasons. Table 5 also shows other underlying differences between sites, including lower MPG in lean sites, possibly reflecting more urban locations and therefore different route structures.

For this reason, we construct a matched subsample of sites that adjust for these underlying

differences. This subsample includes 78 of the 142 sites in the full experimental sample. The excluded sites include nine of the larger sites in the lean group and fifty five of the smaller, more rural sites in the non-lean group. The right side of Table 5 compares the sites across these two groups in this reduced sample and shows that the sites are now statistically indistinguishable across both observable site characteristics and pre-rank posting driver performance.

<< Insert Table 5 about here >>

Appendix Table A2 reproduces the descriptive statistics in Table 2 using the matched sample.

Table 6 reproduces the ITT estimates of Table 3 on the matched sample. The point estimates generally increase and also represent larger percentage standard deviation increases (although they are not statistically different from each other). For example, the Column (2) estimate on Post*Treatment Group 1*Lean shows a 15.1% increase in log gap score, compared to the 13.4% increase estimated in Table 3 Column (2). In general, the results of Table 3 are reproduced and, if anything, strengthened, with the exception of shift score.

<< Insert Table 6 about here >>

Table 7 replicates the fixed effects analysis of Table 4 on the matched subsample. The results are largely reproduced, again with the exception of shift score.

<< Insert Table 7 about here >>

5.4 Additional Analyses

5.4.1 Potential Changes in Underlying Route Characteristics

We also performed three additional analyses to address separate challenges regarding data and experimental design. One question is whether there is anything fundamentally different about the structure or characteristics of the routes driven in lean and non-lean sites that could result in the same observed patterns over this time period. While we believe that this possibility is remote - it would have to affect only Treatment Group 1 (named posting), only at the same six week post-EOBR rollout window as the rank postings and also be orthogonal to the site characteristics on which we based the matched analysis - we perform a placebo test to further rule out this possibility.

For this test, we replace our four outcome variables with "potential" MPG. Potential MPG is the system-calculated variable that response to route characteristics and road and weather conditions, but not to driver performance. Therefore, if any route characteristics changed during this period in the lean named-rank group that led to changes in driver performance, we should observe similar patterns in the potential MPG metric.

The results of this analysis are shown in Appendix Tables A3 (on the full sample) and A4 (on the matched sample). Potential MPG shows no changes during this period, while Actual MPG - which is directly related to driver effort - does. It does not appear, therefore, that underlying changes to the routes are driving the results.

5.4.2 Correcting for Compliance

Coordinating the posting rollouts posed a management challenge for the company, particularly since the rankings were rolled out on a weekly basis across 180 sites in 48 states during the busy winter season. As a consequence, approximately 50% of the sites complied with the rank postings during our experiment. To account for this selective compliance, we followed the approach from [cite] and instrumented actual treatment with assigned treatment. The results of this analysis are shown in Appendix Tables A5 and A6 and are stronger than our earlier analyses (including fixed effects and the matched cohorts). These stronger results are consistent with [choice to comply...]. Further, they provide evidence that our main results likely understate the economic impact of the rank postings on driver performance.

5.4.3 Persistence of effect

Finally, Appendix Tables A7-A10 repeat the analysis shown in Tables 3-6 without restricting the time windows to the 30 days pre- and post-rank postings. These analyses now include 47 days prior and 207 days after the rank postings. We find that, consistent with some attenuation, the standard errors of the estimates are generally larger, but the effect sizes remain essentially unchanged from the narrower-window analysis.

6 How do we know it is "collectivistic orientation" that matters?

Up to this point, we have asserted that the lean intervention created a collectivist-oriented relational contract and that it is this collectivist orientation that drives the different employee response to performance postings. Since relational contracts are, by definition, extremely hard to observe, how do we know that this is the mechanism at work in this experiment?

While we cannot prove this definitively, in this section we provide some supporting evidence for this story. In particular, we present three distinct tests that, taken together, are consistent with our argument. In our first test, we examine the differences in the response between our two treatment groups, the first of which identified the driver performance by name and the second of which identified performance by the anonymous employee ID. In the second test, we look at the effect of rankings on the dispersion of performance within each location. Lastly, we relate driver performance to a proxy measure of "collectivist orientation" based on an employee engagement survey.

Each of these tests relies on research from social psychology to construct predictions that are consistent with the respective roles of individualist and collectivist orientations.

6.1 Named vs IDed Postings

Our reasoning suggests that the positive and negative effect of the named rankings, in both non-Lean and Lean sites respectively, relies on the identifiability of individuals in the rankings. That is, the mechanism we propose for why rankings will have a positive effect in non-lean sites but a negative effect in lean sites is contingent on the rankings personally identifying each of the individuals. The intragroup competition to positively distinguish oneself, which we hypothesize motivates in non-lean sites (where an individualistic orientation prevails), can only unfold if the rankings clearly identify where each individual stands, who is beating whom, who one needs to outdo in order to achieve a higher rank, etc. Similarly, the goal of basking in recognition for good performance and avoiding the shame of being revealed as a poor performer are only relevant in cases where the rankings

personally identify each individual. If rankings were instead posted in an anonymous fashion—one in which individuals could only identify their own performance, but not those of their coworkers—then the incentive of positively distinguishing oneself (which motivates those with an individualistic orientation) would dissipate.

Moreover, anonymous rankings should also undermine the mechanism that we propose for the negative effects of rankings among our Lean sites, where a collectivistic orientation prevails. Concerns about erosion of group cohesion and feeling connected to others should be less likely to arise when rankings are anonymous, since anonymity strips the rankings of their relationally-eroding meaning and significance. That is, since comparisons to known others are not possible when rankings are anonymous, we would expect that the perceived damage to social relations would be attenuated since it is unknown who is beating whom, who is disappointing whom, etc. The adversarial nature of the rankings should be greatly reduced, since one does not know who their adversaries are. As a result, anonymous rankings are less likely to be seen as violating the collectivistic-oriented relational contract that is in place in lean sites.

Overall, anonymity in rankings should strip the rankings of those elements that are central to the positive and negative effects of ranking on performance, respectively, in our non-Lean and Lean sites. If correct, then anonymous rankings should not replicate the pattern of ranking effects that we present above. If incorrect—for instance, if rankings have their effect simply because they convey relative performance feedback, or even more simply because they convey individual performance feedback—then we would expect that anonymous rankings would demonstrate the same pattern of effects as we found for identifiable rankings.

We utilized this distinction between identifiable and anonymous rankings as a means of exploring the validity of our proposed mechanism. In particular, in our study we included a second treatment group—ID postings—which was likewise included in our random assignment of sites to rankings treatment. In this additional treatment condition, performance was posted in an identical manner to that utilized in our named ranking condition except for one critical difference. Rather than identifying employees by name, we identified them on the ranking charts by their employee IDs.

Consistent with the reasoning above, we find no difference between Treatment Group 2 and the

control in our triple-difference analysis so far presented.[2]

6.2 Performance Variance

Another test for our proposed psychological mechanism is to examine performance variance, rather than averages. If the difference in driver response to the postings results from a collective relational contract, we predict that named postings should reduce variance in driver performance within lean sites and raise variance within non-lean sites. We predict this pattern for two reasons. First, employees in highly collectivist environments develop more strongly shared ways of thinking and acting, which should compress performance outcomes once performance information is available. Second, consistency in reactions will likely be more common in cases that involve a direct violation to the group's values and norms, as in the case when named postings are introduced in Lean sites. This is because violations (vs. no violations) are highly salient and lead group members to focus their attention on the nature of the violation. Moreover, when those violations are seen as threats to the predominant norms and culture of the group, identification with the group (i.e., merging one's sense of self up with the group) may increase (Ashforth & Mael, 1992; Tajfel & Turner, 1978) and this in itself can suppress differences among group members and encourage normative responses (Blader & Tyler, 2009; Haslam, 2004).

In sum, our speculation that the effects of Lean are driven by collective orientation would be supported if performance variance among lean/named rank sites is lower than that of other sites. That is, our proposed mechanism would suggest that there should be greater homogeneity of behavior in Lean/named-rank sites. In non-Lean sites, individualistic orientations predominate and thus shared understandings about norms and behavior are diminished. Moreover, in Lean/no-rank sites, there is no inconsistency that might spur a common focus on a given issue (e.g., on a set of behaviors, such as those examined in our study) and there is no perceived threat to draw group members inward to the group and to each other. Under these circumstances, behavior is more susceptible to a wide variety of influences and thus greater variation among individuals is likely.

In contrast, if performance variance is the same between the Lean/named condition and our

[2] Tables 3, 5, 6, 7, A3, A4, A5, A6, A7, A8, A9, A10

other conditions, that would call into question our proposed mechanism. In particular, it could either suggest that a) Lean does not breed a collective orientation and/or b) that the rankings are not interpreted by employees as a violation and threat to the prevailing culture at Lean sites.

Table 8 shows the effect of rank postings on daily performance variance across lean and non-lean sites. Several results are apparent. We find that, overall across both lean and non-lean sites, variance decreases over time. This trend may be due to a learning effect on the part of the drivers or improved instrument calibration. Second, we note that this decreased variance does not occur in non-lean sites with named ranking postings; that is, relative to the control groups, variance increases in non-lean sites with named postings. This may reflect each employee's effort to distinguish themselves from their coworkers. In contrast, in the lean sites, variance either reduces or is statistically no different between the control and Treatment Group 1 (named postings) (in other words, the sum of the coefficients of Post*Treatment Group 1*Lean and Post*Treatment Group 1 is statistically zero).

Lastly, we also observe no effect of IDed rank postings on variance in either the lean or the non-lean sites. In terms of economic magnitudes, the difference in response between lean and non-lean sites ranges from 25% of a standard deviation of log(gap score) variance to 45% of a standard deviation for log(idle time). In sum, these results are consistent with our reasoning that collective orientation will compress performance once performance is revealed, while individualistic orientations will increase dispersion.

<< Insert Table 8 about here >>

6.3 Engagement Survey

Our third test of the proposed mechanism takes the most direct approach, in which we use responses to the company's annual engagement survey. This survey was conducted across 45 sites in July 2014 with 564 driver responses (k= XX). From the survey responses, we created a direct measure of collectivistic orientation as well as a "placebo" measure of individualistic orientation (specifically, instrumental assessments/judgments of the outcomes received from the organization).We now use these two measures to test for whether lean affected a site's orientation and whether that orientation

influenced a driver's response to rankings.

If collectivist orientation is in fact driving our results, we should find the following in the survey data: a) Lean sites should score significantly higher on employee responses on our collectivist orientation index, and b) the collectivist orientation index should produce a similar pattern of results as our lean indicator in our primary triple-differences analysis. Together, these findings would substantiate two of the most critical components of the explanation for our findings, since they would support our arguments that lean does indeed facilitate collective orientation and that collective orientation underlies the rank x lean interaction we find in our focal analysis. In other words, the survey data enables us to more directly and precisely examine (in a subset of our data) whether the factor that is most central to our proposed mechanism relates, in accordance with our predictions, to both its antecedent (Lean) and its consequences for performance. In addition to testing these specific relationships, we also conducted the same analyses using the instrumentality index. Doing so provides a counterpoint for interpreting our findings, enabling us to determine if a) instrumental concerns, rather than collectivistic orientation, better explain our effects and b) if a factor common to both indices—e.g., overall satisfaction—is the factor that actually underlies our findings.

Table 9 shows the relationships between lean sites and our collective and instrumental measures (defined above). This table shows that lean sites are associated with higher scores for both measures.

<< Insert Table 9 about here >>

Since this is a cross-sectional analysis, we may be concerned that underlying differences between lean and non-lean sites drive the differences engagement survey responses. Accordingly, Table 10 shows the same analysis as the previous table using a sub-sample of survey responses in which lean and non-lean sites were matched by size, region, and driver race, age and tenure. The results attenuate somewhat but are statistically the same as the unmatched sample. In particular, the association between lean and collective engagement remains robust.

<< Insert Table 10 about here >>

Appendix Tables A11 and A12 show the disaggregated categories that represent the four distinct psychological constructs used to construct the overall collectivist orientation index. The coefficient

estimates on lean are all positive, although with varying levels of statistical significance. These two tables together show that lean is associated with higher engagement overall.

Table 11 shows the main result of this analysis: a higher score on the collectivist index is associated with a more negative response to named postings, while no such response is observed using the instrumental (placebo) index. This result is shown graphically in Figure 3, which divides the collective and instrumental indices into deciles and plots each cohort's response to the named postings, relative to the lowest decile. We can see from this figure an increasingly negative response to named posting as collective decile increase, while no such pattern is observed for across the instrumental deciles.

<< Insert Table 11 about here >>

<< Insert Figure 3 about here >>

One cautionary note about this analysis is that there are appears to be some pre-treatment differences between drivers with high collective values in Treatment Group 1 and the control group (see the Treatment Group 1*[Category] coefficient in Table 11), so we cannot rule out definitively that our observed patterns are not driven by underlying differences between these two groups. Mitigating this concern is that Table A15 shows that the magnitude of the treatment response appears unrelated to the degree of pre-treatment differences between these two groups. However, these results must be interpreted with caution.

From this last test, our results indicate: a) a pattern for the collective orientation index that is highly consistent with and supportive of our proposed mechanism, b) a pattern for the instrumental index, an index tied more closely to individualistic orientation, that is not consistent with and thus not likely to explain our primary findings, and thus c) suggestive evidence that it is not something common to both indices—e.g., a generalized sense of satisfaction—that drives our results but rather something unique to collective orientation.

The results of all three tests together - while not definitive - are supportive of the explanation that lean affects collectivist orientations, which in turn influence the drivers' responses to the rank postings. Any other proposed mechanism would also have to also explain these three observed patterns, each derived from a very different analysis.

7 Conclusion

In this study, we randomized the posting of employee performance rankings across a company that was midway through a costly, multi-year process of altering its relational contract with its employees. Employees working in locations with the original contract responded positively to the performance postings, with their performance improving 2-4% relative to the control group (depending on the performance measure). In contrast, employees in the sites with the newer contract responded negatively, with their performance declining 4-13% relative to the control group.

It appears that these different responses are driven by the individualistic orientation of the initial contract and the collectivist orientation of the new contract. This new contract is based on the Toyota Production System which emphasizes the value of teamwork and cooperation, as well as the subservient role of management whose primary task is to enable the front line workers. Research in social psychology has found that employees respond poorly to perceived inconsistencies in leaders' messages [rephrase]. Our findings support this result, with the posting of individualistic performance rankings representing a violation of the collectivist contract rolled out by the company.

We have several directions for future research. The timing of this study did not permit us to simultaneously randomize the implementation of both the management practice and the relational contract, since the practice had to be rolled out over the period of three months at the end of 2013 and the initiative to alter the firm's relational contract began in 2011 and required a minimum of five years to be implemented fully. To accommodate this timing mismatch, we stratified the randomized ranking postings by the lean status of each location. In later studies, we aim to randomize the rollout of the relational contract itself, a much lengthier and complex process that should allow us to make more definitive statements about the direct impact of relational contracts on employee productivity. Aside from this direct effect, a further area to explore is process of altering these relational contracts themselves. Specifically, we would like to understand the factors that determine differences in adoption success. Finally, beyond single firm studies, we would like to extend this research across firms, industries and geographies.

8 References

Figures

Figure 1: Impact of Rankings on Driver Performance

Figure 1a: Median effect with each site weighted equally

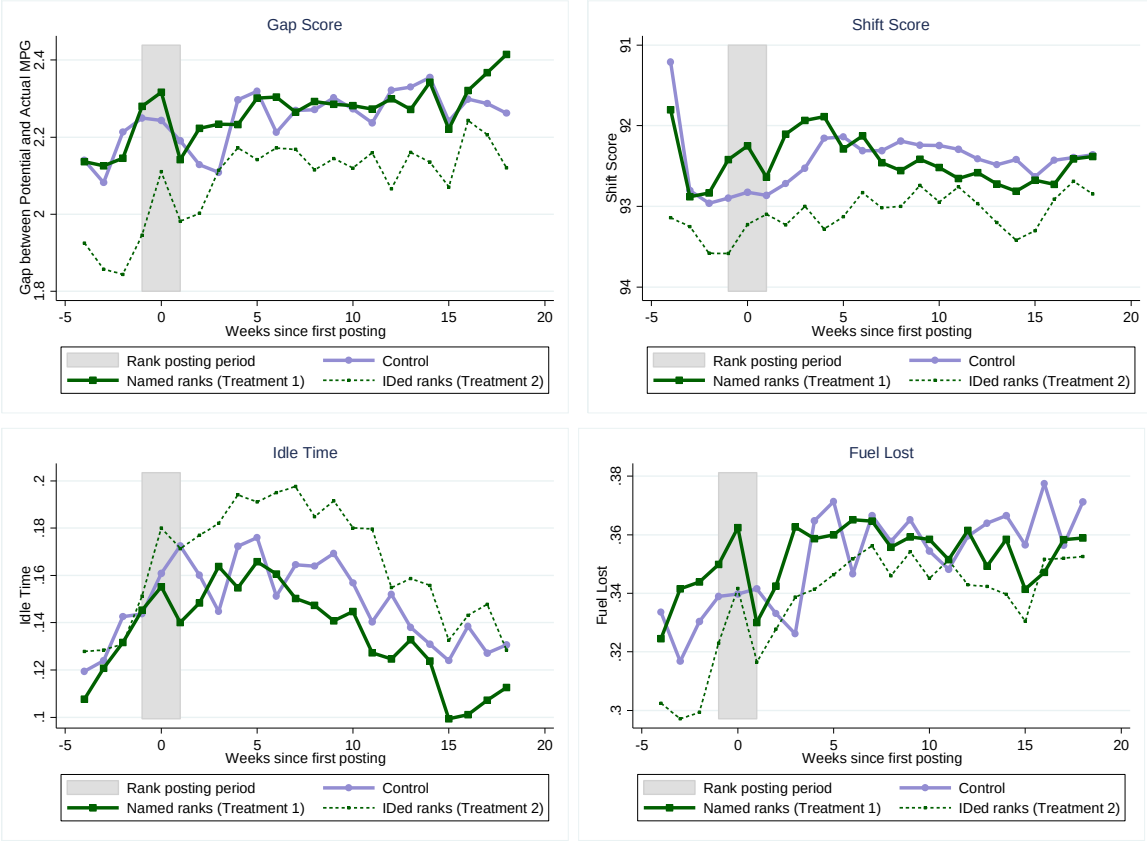


Figure 1b: Average affect with each driver weighted equally

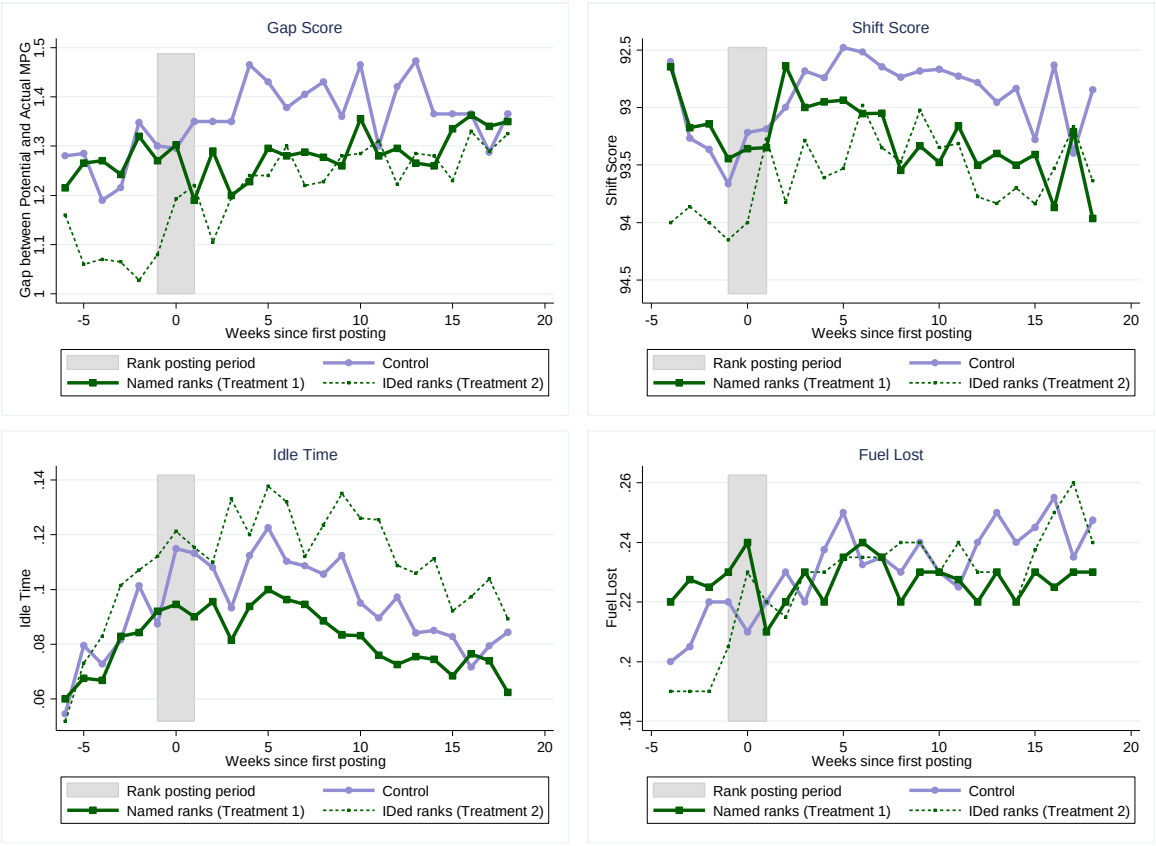


Figure 2: Impact of Rankings by Site Type

Figure 2a: Median effect with each site weighted equally

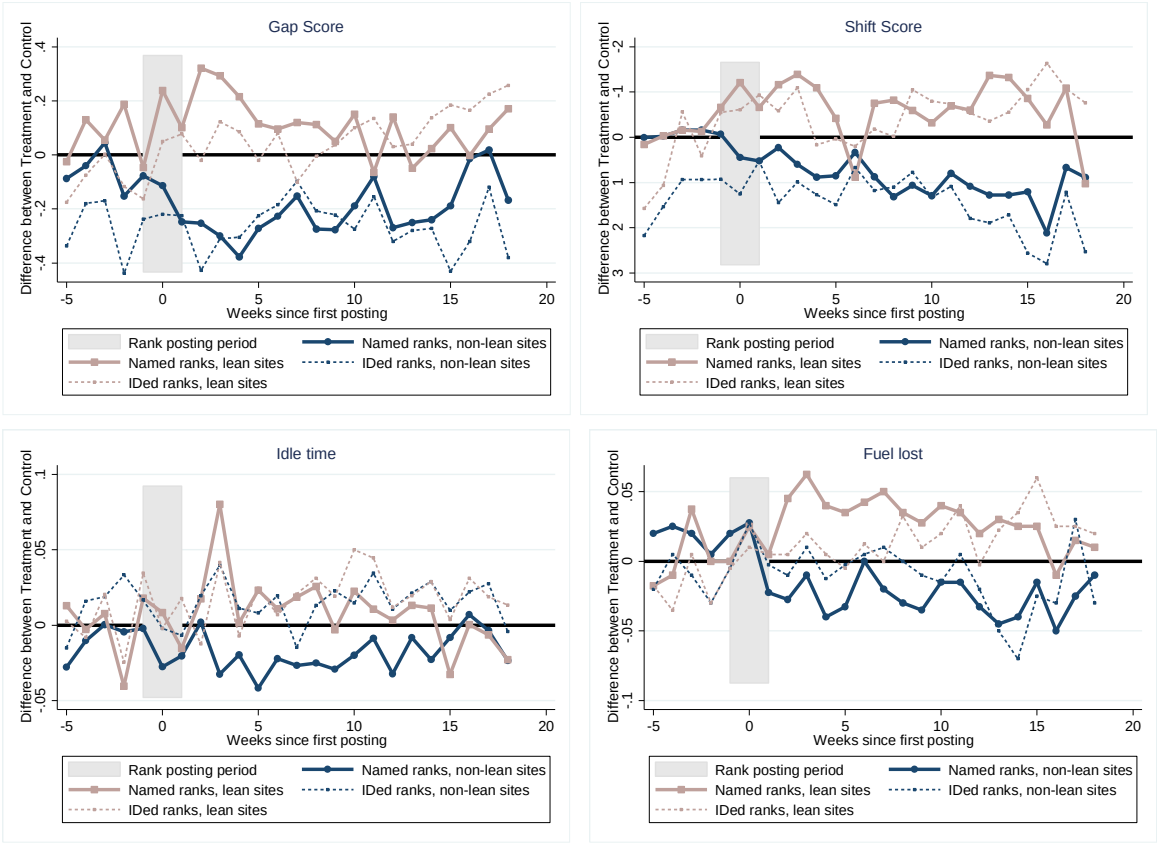


Figure 2b: Average affect with each driver weighted equally

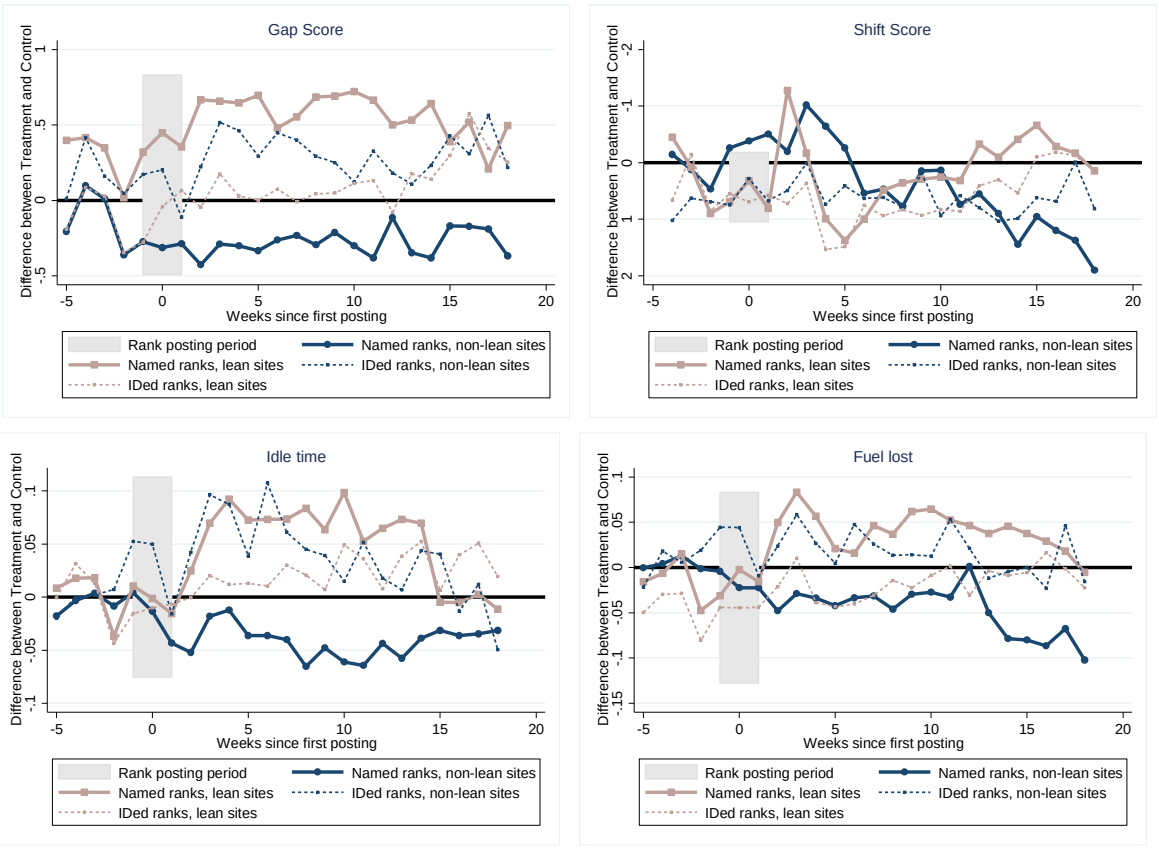
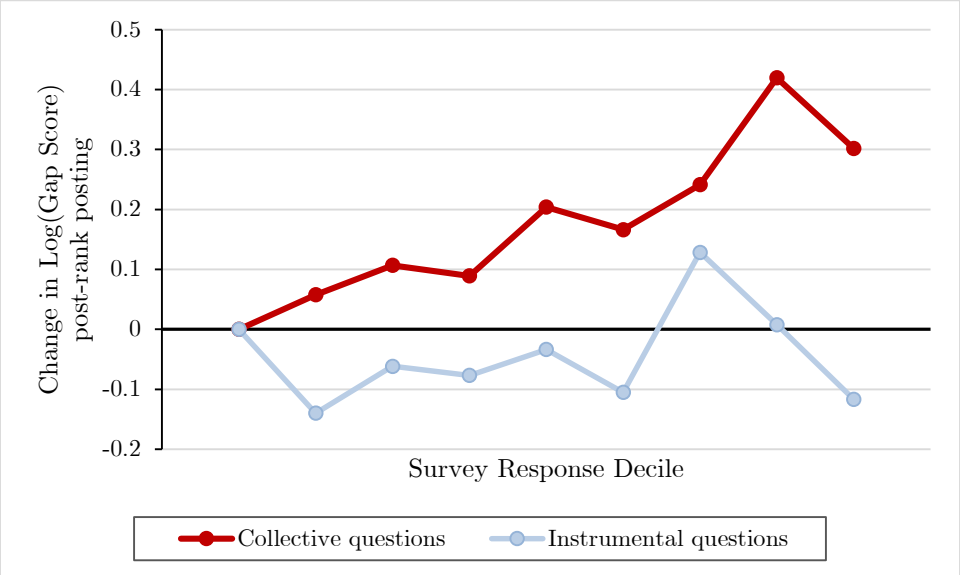
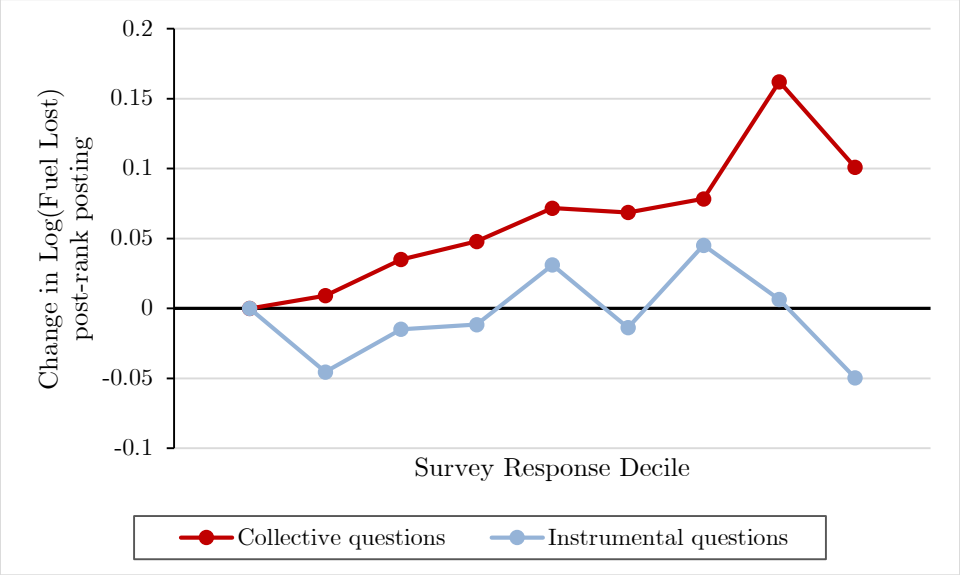


Figure 3: Rank posting response by engagement survey response





Tables

Table 1: Descriptive statistics

	All Sites				Sample				Control	Treatment 1 (names)	Diff	Treatment 2 (IDs)	Diff
	Mean	Med	Min	Max	Mean	Med	Min	Max	Mean	Mean	p-value	Mean	p-value
Site characteristics													
# sites	275	n/a	n/a	n/a	142	n/a	n/a	n/a	47	50	n/a	45	n/a
Lean status	0.37	n/a	0	1	0.34	n/a	0	1	0.30	0.26	0.681	0.47	0.098
Tractors / site	30.96	22	5	151	24.71	21	5	87	25	25.32	0.924	23.73	0.664
Distance / trip	127.64	127.42	43.76	196.46	128.65	128.5	56.08	200.65	124.08	130.63	0.309	131.24	0.247
Pre-ranking driver performance (site mean)													
Miles per gallon	6.81	6.83	5.35	8.23	6.82	6.80	5.66	8.37	6.76	6.88	0.247	6.82	0.558
Gap score	2.19	2.1	0.57	6.99	2.10	1.94	0.78	6.43	2.18	2.14	0.787	1.98	0.310
Shift score	90.82	91.2	74.78	97.28	91.07	91.53	74.43	97.41	90.77	90.69	0.902	91.79	0.149
Excess idle time	0.12	0.1	0	0.72	0.13	0.11	0.02	0.72	0.12	0.12	0.838	0.14	0.429
Fuel lost	0.33	0.32	0.12	0.74	0.33	0.33	0.14	0.72	0.34	0.35	0.722	0.31	0.185

Table 2: Effect of rankings on all sites

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1	0.0061	0.0107	-0.6105	-0.6979*	0.0004	0.0007	-0.0007	0.0029
	(0.0213)	(0.0206)	(0.3902)	(0.3736)	(0.0076)	(0.0072)	(0.0077)	(0.0073)
Post*Treatment group 2	0.0429	0.0490**	-0.3728	-0.4428	0.0138*	0.0124*	0.0118	0.0128*
	(0.0262)	(0.0221)	(0.3786)	(0.3622)	(0.0077)	(0.0074)	(0.0082)	(0.0076)
Post	0.0139	0.0061	0.1084	0.2289	0.0216***	0.0188***	0.0079	0.0026
	(0.0168)	(0.0149)	(0.3121)	(0.3016)	(0.0047)	(0.0046)	(0.0054)	(0.0051)
Treatment group 1 (names)	-0.0030	-0.0096	0.1024	0.0324	-0.0048	-0.0116	0.0081	-0.0020
	(0.0445)	(0.0262)	(0.4995)	(0.3689)	(0.0088)	(0.0077)	(0.0118)	(0.0099)
Treatment group 2 (IDs)	-0.0845*	-0.0147	0.8754*	0.0980	0.0008	0.0110	-0.0161	-0.0043
	(0.0477)	(0.0363)	(0.4861)	(0.4372)	(0.0085)	(0.0072)	(0.0129)	(0.0142)
Constant	0.9366***	1.7845***	92.5090***	90.4229***	0.1001***	0.2794***	0.2524***	-0.0842*
	(0.0327)	(0.1247)	(0.3352)	(1.4176)	(0.0056)	(0.0366)	(0.0080)	(0.0476)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.003	0.128	0.003	0.044	0.005	0.050	0.002	0.102

Table 3: Effect of rankings on lean and non-lean sites

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1358***	0.1338***	-1.8625**	-1.9657***	0.0394***	0.0352***	0.0593***	0.0538***
	(0.0384)	(0.0383)	(0.7702)	(0.7120)	(0.0136)	(0.0131)	(0.0144)	(0.0127)
Post*Treatment group 2*Lean	0.0251	0.0329	-0.4615	-0.6557	-0.0080	-0.0073	0.0093	0.0131
	(0.0522)	(0.0462)	(0.7701)	(0.7176)	(0.0159)	(0.0152)	(0.0168)	(0.0161)
Post*Treatment group 1	-0.0423*	-0.0375	0.0904	0.0366	-0.0130	-0.0113	-0.0217***	-0.0164**
	(0.0251)	(0.0232)	(0.4018)	(0.3800)	(0.0084)	(0.0080)	(0.0077)	(0.0080)
Post*Treatment group 2	0.0329	0.0333	-0.2777	-0.2295	0.0178	0.0160	0.0079	0.0064
	(0.0392)	(0.0336)	(0.3835)	(0.3777)	(0.0134)	(0.0128)	(0.0116)	(0.0117)
Post*Lean	-0.0303	-0.0269	0.9783	1.0118*	0.0034	0.0027	-0.0098	-0.0090
	(0.0326)	(0.0295)	(0.6547)	(0.6043)	(0.0091)	(0.0090)	(0.0111)	(0.0100)
Treatment group 1*Lean	0.1434*	0.0415	0.5095	0.7454	0.0066	-0.0166	0.0149	0.0066
	(0.0864)	(0.0658)	(0.9945)	(0.7343)	(0.0180)	(0.0163)	(0.0242)	(0.0241)
Treatment group 2*Lean	0.0549	0.0814	-0.7479	-0.3329	-0.0002	0.0141	-0.0022	0.0294
	(0.0895)	(0.0682)	(0.9417)	(0.7676)	(0.0165)	(0.0125)	(0.0250)	(0.0266)
Post	0.0263	0.0180	-0.2856	-0.1838	0.0203***	0.0178***	0.0118*	0.0066
	(0.0222)	(0.0186)	(0.2532)	(0.2425)	(0.0064)	(0.0063)	(0.0065)	(0.0065)
Lean	-0.0937	-0.0814*	0.1299	0.2731	-0.0014	-0.0123	-0.0139	-0.0264
	(0.0590)	(0.0445)	(0.6939)	(0.5931)	(0.0106)	(0.0093)	(0.0160)	(0.0169)
Treatment group 1 (names)	-0.0578	-0.0296	-0.0684	-0.1717	-0.0072	-0.0071	0.0022	-0.0062
	(0.0575)	(0.0337)	(0.6552)	(0.4627)	(0.0116)	(0.0102)	(0.0146)	(0.0131)
Treatment group 2 (IDs)	-0.0996	-0.0503	1.2832**	0.2648	0.0012	0.0049	-0.0124	-0.0174
	(0.0659)	(0.0426)	(0.5405)	(0.5338)	(0.0121)	(0.0104)	(0.0162)	(0.0166)
Constant	0.9741***	1.7811***	92.4570***	90.7710***	0.1007***	0.2767***	0.2579***	-0.0854*
	(0.0453)	(0.1292)	(0.4212)	(1.4685)	(0.0081)	(0.0371)	(0.0098)	(0.0488)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.009	0.130	0.004	0.045	0.007	0.051	0.004	0.104

Notes: SE clustered by SIC, window (5,30), controls include XYZ, winsorization

Table 4: Analysis using date and driver fixed effects

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1026***	0.1067***	-1.9433***	-1.9459***	0.0279**	0.0254*	0.0501***	0.0441***
	(0.0359)	(0.0355)	(0.6079)	(0.6041)	(0.0140)	(0.0132)	(0.0141)	(0.0124)
Post*Treatment group 2*Lean	0.0544	0.0561	-1.2449**	-1.2635**	-0.0087	-0.0069	0.0169	0.0192
	(0.0396)	(0.0389)	(0.5879)	(0.5845)	(0.0157)	(0.0152)	(0.0143)	(0.0136)
Post*Treatment group 1	-0.0424*	-0.0452**	0.3998	0.3974	-0.0133	-0.0109	-0.0231***	-0.0180**
	(0.0229)	(0.0221)	(0.3492)	(0.3484)	(0.0087)	(0.0084)	(0.0079)	(0.0077)
Post*Treatment group 2	0.0112	0.0115	0.0615	0.0674	0.0143	0.0131	0.0042	0.0021
	(0.0304)	(0.0292)	(0.3261)	(0.3252)	(0.0127)	(0.0126)	(0.0099)	(0.0104)
Post*Lean	-0.0368	-0.0415	1.2533**	1.2798***	0.0062	0.0052	-0.0130	-0.0126
	(0.0272)	(0.0271)	(0.4947)	(0.4896)	(0.0094)	(0.0090)	(0.0105)	(0.0092)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	-0.0049	-0.0027	-0.8095***	-0.8222***	-0.0069	-0.0064	-0.0066	-0.0067
	(0.0216)	(0.0208)	(0.2753)	(0.2741)	(0.0081)	(0.0080)	(0.0082)	(0.0080)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9495***	1.8575***	92.9190***	88.6765***	0.1260***	0.1821***	0.2564***	-0.0576***
	(0.0119)	(0.0484)	(0.1367)	(0.5761)	(0.0057)	(0.0131)	(0.0052)	(0.0195)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.546	0.559	0.601	0.602	0.285	0.298	0.459	0.509

Table 5: Comparison of lean and non-lean sites

	Full sample			Matched sample		
	Non-lean Mean	Lean Mean	Diff p-value	Non-lean Mean	Lean Mean	Diff *p*-value
# sites	94	48	n/a	39	39	n/a
Tractors / site	20.35	33.25	0.000	25.95	27.51	0.581
Distance / trip	128.04	127.53	0.609	128.04	127.53	0.937
Miles per gallon	6.90	6.72	0.039	6.76	6.71	0.602
Gap score	2.14	2.04	0.537	2.00	2.03	0.838
Shift score	90.35	91.55	0.076	91.62	91.66	0.950
Excess idle time	0.12	0.13	0.781	0.12	0.13	0.815
Fuel lost	0.34	0.33	0.473	0.32	0.33	0.753
Eastern region	0.27	0.39	0.155	0.37	0.38	0.865
Central region	0.41	0.37	0.626	0.44	0.38	0.626
Western region	0.32	0.24	0.357	0.20	0.23	0.701
Control group	0.35	0.29	0.480	0.39	0.27	0.245
Treatment group 1	0.39	0.27	0.149	0.32	0.24	0.467
Treatment group 2	0.26	0.44	0.027	0.29	0.49	0.072

Table 6: Matched analysis

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1485***	0.1510***	0.3944	0.0457	0.0523***	0.0522***	0.0585***	0.0582***
	(0.0493)	(0.0433)	(0.7437)	(0.7280)	(0.0172)	(0.0157)	(0.0174)	(0.0153)
Post*Treatment group 2*Lean	0.0034	0.0254	0.7460	0.4590	-0.0203	-0.0151	0.0067	0.0084
	(0.0626)	(0.0532)	(0.6049)	(0.5673)	(0.0192)	(0.0181)	(0.0196)	(0.0193)
Post*Treatment group 1	-0.0623*	-0.0496	-0.7339	-0.6517	-0.0214*	-0.0213*	-0.0238**	-0.0197*
	(0.0343)	(0.0327)	(0.5263)	(0.5547)	(0.0123)	(0.0115)	(0.0110)	(0.0112)
Post*Treatment group 2	0.0350	0.0387	-0.3744	-0.3583	0.0298*	0.0258	0.0073	0.0094
	(0.0503)	(0.0443)	(0.3742)	(0.4068)	(0.0173)	(0.0166)	(0.0149)	(0.0153)
Post*Lean	-0.0435	-0.0425	-0.3965	-0.2911	-0.0017	-0.0050	-0.0139	-0.0126
	(0.0408)	(0.0316)	(0.4778)	(0.4521)	(0.0113)	(0.0106)	(0.0126)	(0.0122)
Treatment group 1*Lean	0.0582	0.0384	0.3169	-0.8576	-0.0011	0.0015	-0.0131	0.0155
	(0.1096)	(0.0934)	(1.0673)	(0.9340)	(0.0253)	(0.0254)	(0.0268)	(0.0328)
Treatment group 2*Lean	-0.0848	0.0842	-0.2996	-1.4198	-0.0109	0.0208	-0.0400	0.0379
	(0.1134)	(0.0825)	(1.0487)	(0.8966)	(0.0202)	(0.0173)	(0.0295)	(0.0306)
Post	0.0496*	0.0347	-0.1958	-0.0739	0.0245**	0.0216**	0.0173**	0.0113
	(0.0296)	(0.0250)	(0.2566)	(0.3057)	(0.0095)	(0.0094)	(0.0082)	(0.0086)
Lean	0.0135	-0.0331	-0.3308	0.7914	0.0022	-0.0156	0.0198	-0.0179
	(0.0839)	(0.0649)	(0.7787)	(0.7736)	(0.0138)	(0.0139)	(0.0189)	(0.0246)
Treatment group 1 (names)	-0.0242	0.0064	0.0517	0.3504	-0.0003	-0.0074	0.0022	0.0078
	(0.0712)	(0.0628)	(0.7174)	(0.6182)	(0.0174)	(0.0192)	(0.0177)	(0.0240)
Treatment group 2 (IDs)	0.0396	-0.0702	0.7376	0.8761	0.0094	0.0044	0.0105	-0.0263
	(0.0813)	(0.0680)	(0.6374)	(0.8571)	(0.0150)	(0.0142)	(0.0207)	(0.0261)
Constant	0.8846***	1.7351***	93.3009***	89.4941***	0.0973***	0.2914***	0.2427***	-0.0876
	(0.0574)	(0.1485)	(0.4419)	(1.6906)	(0.0100)	(0.0356)	(0.0126)	(0.0538)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	60002	60002	60002	60002	60002	60002	60002	60002
Adjusted R-squared	0.007	0.127	0.003	0.050	0.009	0.056	0.004	0.102

Table 7: Matched analysis with date and driver fixed effects

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1088***	0.1092***	-0.1514	-0.1495	0.0360**	0.0350**	0.0453***	0.0435***
	(0.0392)	(0.0380)	(0.6861)	(0.6841)	(0.0172)	(0.0161)	(0.0152)	(0.0135)
Post*Treatment group 2*Lean	0.0118	0.0214	-0.0348	-0.0812	-0.0267	-0.0234	0.0031	0.0043
	(0.0436)	(0.0426)	(0.5283)	(0.5263)	(0.0179)	(0.0175)	(0.0163)	(0.0158)
Post*Treatment group 1	-0.0525*	-0.0514*	-0.4054	-0.4153	-0.0215*	-0.0196*	-0.0231**	-0.0204*
	(0.0308)	(0.0303)	(0.4953)	(0.4945)	(0.0114)	(0.0110)	(0.0114)	(0.0107)
Post*Treatment group 2	0.0317	0.0274	-0.1981	-0.1755	0.0280*	0.0259*	0.0096	0.0080
	(0.0397)	(0.0388)	(0.3252)	(0.3275)	(0.0152)	(0.0152)	(0.0137)	(0.0141)
Post*Lean	-0.0299	-0.0355	0.0595	0.0869	0.0026	0.0006	-0.0078	-0.0086
	(0.0279)	(0.0268)	(0.4627)	(0.4589)	(0.0112)	(0.0109)	(0.0110)	(0.0102)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	0.0065	0.0106	-0.2165	-0.2353	-0.0070	-0.0059	-0.0018	-0.0019
	(0.0269)	(0.0259)	(0.3157)	(0.3164)	(0.0113)	(0.0110)	(0.0105)	(0.0102)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9295***	1.8671***	93.4125***	89.6393***	0.1297***	0.2003***	0.2558***	-0.0455*
	(0.0161)	(0.0640)	(0.1689)	(0.6491)	(0.0073)	(0.0155)	(0.0072)	(0.0250)
Controls	N	Y	N	Y	N	Y	N	Y
Date and Driver Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	60002	60002	60002	60002	60002	60002	60002	60002
Adjusted R-squared	0.559	0.572	0.620	0.621	0.289	0.303	0.470	0.521

Table 8: Effect on variance

Coefficient of variation:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	-0.0427*	-0.0436**	0.0044	0.0044	-0.2466***	-0.2466***	-0.0184	-0.0182
	(0.0239)	(0.0205)	(0.0053)	(0.0047)	(0.0787)	(0.0714)	(0.0354)	(0.0282)
Post*Treatment group 2*Lean	0.0042	-0.0102	0.0003	0.0001	-0.0484	-0.0945	0.0265	0.0025
	(0.0241)	(0.0214)	(0.0050)	(0.0043)	(0.0776)	(0.0663)	(0.0339)	(0.0289)
Post*Treatment group 1	0.0389***	0.0325***	-0.0019	-0.0015	0.1335***	0.1124***	0.0311*	0.0251
	(0.0129)	(0.0117)	(0.0029)	(0.0027)	(0.0421)	(0.0365)	(0.0177)	(0.0158)
Post*Treatment group 2	-0.0169	-0.0097	-0.0011	-0.0012	-0.0171	0.0161	-0.0420*	-0.0286
	(0.0166)	(0.0145)	(0.0029)	(0.0027)	(0.0516)	(0.0413)	(0.0231)	(0.0196)
Post*Lean	-0.0059	-0.0010	-0.0032	-0.0027	0.0867	0.1178**	-0.0037	0.0026
	(0.0155)	(0.0142)	(0.0036)	(0.0031)	(0.0543)	(0.0479)	(0.0230)	(0.0197)
Treatment group 1*Lean	-0.0034	-0.0024	0.0032	0.0056	-0.0029	0.0073	0.0038	-0.0208
	(0.0171)	(0.0155)	(0.0036)	(0.0034)	(0.0591)	(0.0556)	(0.0249)	(0.0210)
Treatment group 2*Lean	-0.0573***	-0.0158	0.0076**	0.0146***	-0.1694***	-0.1565***	-0.0463*	0.0145
	(0.0173)	(0.0174)	(0.0034)	(0.0033)	(0.0581)	(0.0550)	(0.0244)	(0.0235)
Post	-0.0253***	-0.0199**	0.0070***	0.0064***	-0.2051***	-0.1862***	-0.0180	-0.0154
	(0.0087)	(0.0079)	(0.0018)	(0.0017)	(0.0289)	(0.0252)	(0.0120)	(0.0108)
Lean	0.0493***	0.0175	-0.0023	-0.0105***	0.1039**	0.0505	0.0447***	0.0120
	(0.0111)	(0.0120)	(0.0026)	(0.0025)	(0.0405)	(0.0423)	(0.0162)	(0.0164)
Treatment group 1 (names)	0.0149	0.0067	0.0040**	0.0024	-0.0476	-0.0385	-0.0044	0.0066
	(0.0096)	(0.0091)	(0.0020)	(0.0020)	(0.0317)	(0.0293)	(0.0130)	(0.0122)
Treatment group 2 (IDs)	0.0729***	0.0326***	-0.0091***	-0.0078***	0.0732*	0.0084	0.0463***	0.0019
	(0.0122)	(0.0121)	(0.0020)	(0.0022)	(0.0393)	(0.0351)	(0.0172)	(0.0163)
Constant	0.5712***	0.3053***	0.0733***	0.1375***	1.6574***	-0.0409	0.7965***	1.1360***
	(0.0066)	(0.0695)	(0.0012)	(0.0152)	(0.0224)	(0.1894)	(0.0086)	(0.0944)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	5343	5343	5343	5343	5285	5285	5343	5343
Adjusted R-squared	0.024	0.240	0.022	0.192	0.029	0.284	0.007	0.280

Table 9: Lean and employee engagement

Dependent variable:	Collective questions		Instrumental questions	
	(1)	(2)	(3)	(4)
Lean	0.2268	0.2735**	0.2954*	0.3314**
	(0.1606)	(0.1234)	(0.1753)	(0.1607)
Constant	3.3001***	3.2945***	3.2542***	2.7641***
	(0.1681)	(0.4946)	(0.1697)	(0.4358)
Demographic Controls	No	Yes	No	Yes
Observations	561	561	564	564
Adjusted R-squared	0.026	0.127	0.016	0.075

Table 10: Matched analysis of Lean and employee engagement

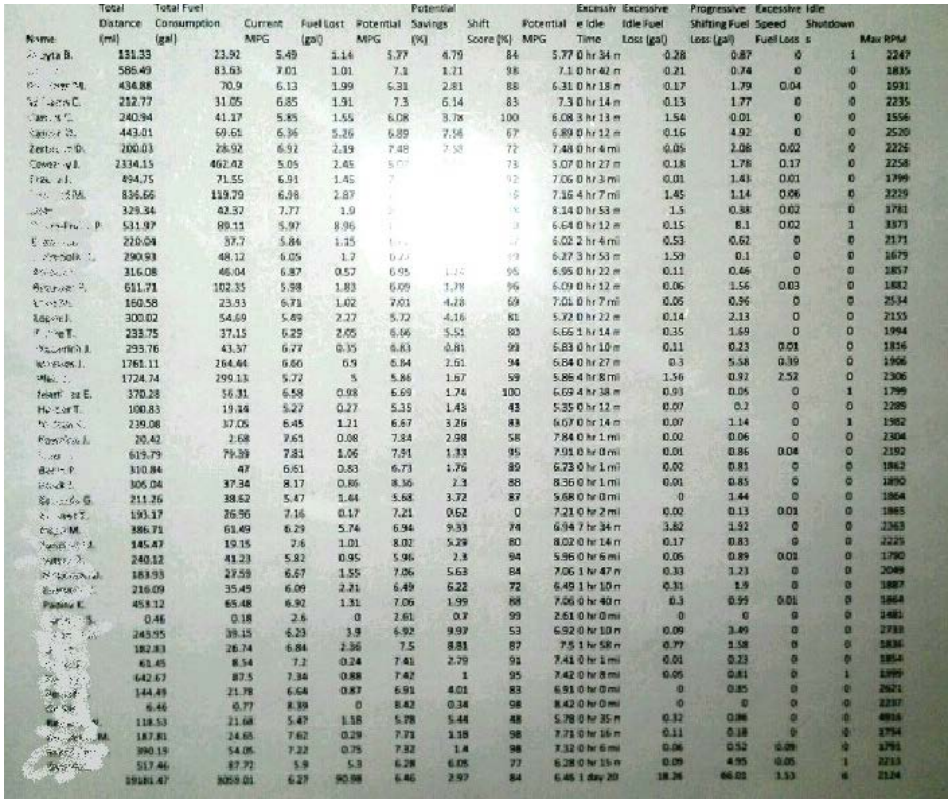
Dependent variable:	Collective questions		Instrumental questions	
	(1)	(2)	(3)	(4)
Lean	0.1731	0.2736**	0.1730	0.2353
	(0.1577)	(0.1298)	(0.1978)	(0.1965)
Constant	3.3550***	3.3327***	3.3138***	2.9711***
	(0.2183)	(0.3734)	(0.1881)	(0.6146)
Demographic Controls	No	Yes	No	Yes
Observations	396	396	399	399
Adjusted R-squared	0.029	0.128	0.009	0.056

Table 11: Effect of ranking and engagement on driver performance

Dependent variable:	Log(Gap Score)			
Category:	Collective questions		Instrumental questions	
	(1)	(2)	(3)	(4)
Post*Treatment group 1*[Category]	0.1151***	0.1182**	0.0407	0.0493
	(0.0423)	(0.0526)	(0.0368)	(0.0488)
Post*Treatment group 2*[Category]	0.0322	0.0042	-0.0243	-0.0517
	(0.0629)	(0.0815)	(0.0574)	(0.1012)
Post*Treatment group 1	-0.4082***	-0.4440***	-0.1741	-0.2261
	(0.1273)	(0.1531)	(0.1157)	(0.1516)
Post*Treatment group 2	-0.0511	-0.0500	0.1370	0.1377
	(0.2269)	(0.3196)	(0.2195)	(0.4129)
Post*[Category]	-0.0628*	-0.0554	-0.0196	-0.0248
	(0.0339)	(0.0455)	(0.0264)	(0.0391)
Treatment group 1*[Category]	0.1151*	0.1384**	0.0876	0.1237
	(0.0600)	(0.0656)	(0.0671)	(0.0749)
Treatment group 2*[Category]	-0.0578	0.0241	-0.0544	-0.0516
	(0.0677)	(0.0816)	(0.0596)	(0.0835)
Post	0.2570**	0.2497*	0.1246	0.1599
	(0.1027)	(0.1316)	(0.0914)	(0.1224)
[Category]	-0.0309	-0.0515	-0.0404	-0.0416
	(0.0397)	(0.0405)	(0.0461)	(0.0485)
Treatment group 1 (Names)	-0.3260	-0.3650	-0.2278	-0.3003
	(0.2179)	(0.2294)	(0.2523)	(0.2664)
Treatment group 2 (IDs)	0.2767	0.0437	0.2623	0.3041
	(0.2145)	(0.2581)	(0.1981)	(0.2679)
Constant	2.1358***	2.3431***	2.0377***	2.1163***
	(0.3214)	(0.3606)	(0.2985)	(0.3487)
Sample	Full	Matched	Full	Matched
Observations	35187	26065	35385	26263
Adjusted R-squared	0.106	0.117	0.095	0.112

Appendix Figures and Tables

Figure A1: Sample Rank Posting



Name	Total Distance (mi)	Total Fuel Consumption (gal)	Current MPG	Fuel Lost (gal)	Potential MPG	Potential Savings (%)	Shift Score (%)	Potential MPG	Excessiv e Idle Time	Excessive Idle Fuel Loss (gal)	Progressive Shifting Fuel Loss (gal)	Excessive Speed Fuel Loss	Idle Shutdown s	Max RPM
[redacted]	131.33	23.92	5.49	1.14	5.77	4.79	84	5.77	0 hr 34 m	0.28	0.87	0	1	2247
[redacted]	586.49	83.63	7.01	1.01	7.1	1.21	98	7.1	0 hr 42 m	0.21	0.74	0	0	1835
[redacted]	434.88	70.9	6.13	1.99	6.31	2.81	88	6.31	0 hr 18 m	0.17	1.79	0.04	0	1931
[redacted]	212.77	31.05	6.85	1.91	7.3	6.14	83	7.3	0 hr 14 m	0.13	1.77	0	0	2235
[redacted]	240.94	41.17	5.85	1.55	6.08	3.78	100	6.08	3 hr 13 m	1.54	0.01	0	0	1556
[redacted]	443.01	69.61	6.36	5.26	6.89	7.56	67	6.89	0 hr 12 m	0.16	4.92	0	0	2520
[redacted]	200.03	28.92	6.92	2.19	7.48	7.58	72	7.48	0 hr 4 mi	0.05	2.06	0.02	0	2226
[redacted]	2334.15	462.42	5.05	2.45	5.07		73	5.07	0 hr 27 m	0.18	1.76	0.17	0	2258
[redacted]	494.75	71.55	6.91	1.45	7		92	7.06	0 hr 3 mi	0.01	1.43	0.01	0	1799
[redacted]	836.66	119.79	6.98	2.87	[redacted]		76	7.16	4 hr 7 mi	1.45	1.14	0.06	0	2229
[redacted]	329.34	42.37	7.77	1.9	[redacted]		93	8.14	0 hr 53 m	1.5	0.38	0.02	0	1781
[redacted]	531.97	89.11	5.97	8.96	[redacted]		3	6.64	0 hr 12 m	0.15	8.1	0.02	1	1973
[redacted]	220.04	37.7	5.84	1.15	[redacted]		97	6.02	2 hr 4 mi	0.53	0.62	0	0	2171
[redacted]	290.93	48.12	6.05	1.7	6.23		93	6.27	3 hr 53 m	1.59	0.1	0	0	1679
[redacted]	316.08	46.04	6.87	0.57	6.95	1.16	96	6.95	0 hr 22 m	0.11	0.46	0	0	1857
[redacted]	611.71	102.35	5.98	1.83	6.09	1.79	96	6.09	0 hr 12 m	0.06	1.56	0.03	0	1882
[redacted]	160.58	23.93	6.71	1.02	7.01	4.28	69	7.01	0 hr 7 mi	0.05	0.96	0	0	2534
[redacted]	300.02	54.69	5.49	2.27	5.72	4.16	61	5.72	0 hr 22 m	0.14	2.13	0	0	2155
[redacted]	233.75	37.15	6.29	2.65	6.66	5.55	83	6.66	1 hr 14 m	0.35	1.69	0	0	1994
[redacted]	293.76	43.37	6.77	0.35	6.83	0.81	99	6.83	0 hr 10 m	0.11	0.23	0.01	0	1816
[redacted]	1761.11	264.44	6.66	0.9	6.84	2.61	94	6.84	0 hr 27 m	0.3	5.58	0.39	0	1906
[redacted]	1724.74	299.13	5.72	5	5.86	1.67	59	5.86	4 hr 8 mi	1.56	0.92	2.52	0	2306
[redacted]	370.28	56.31	6.58	0.98	6.69	1.74	100	6.69	4 hr 38 m	0.93	0.06	0	1	1799
[redacted]	100.83	19.34	5.22	0.27	5.35	1.43	43	5.35	0 hr 12 m	0.07	0.2	0	0	2289
[redacted]	239.08	37.05	6.45	1.21	6.67	3.26	83	6.67	0 hr 14 m	0.07	1.14	0	1	1982
[redacted]	20.42	2.68	7.61	0.08	7.84	2.98	58	7.84	0 hr 1 mi	0.02	0.06	0	0	2304
[redacted]	619.79	79.39	7.81	1.06	7.91	1.33	95	7.91	0 hr 0 mi	0.01	0.86	0.04	0	2192
[redacted]	310.84	47	6.61	0.83	6.73	1.76	89	6.73	0 hr 1 mi	0.02	0.81	0	0	1862
[redacted]	306.04	37.34	8.17	0.86	8.36	2.3	88	8.36	0 hr 1 mi	0.01	0.85	0	0	1890
[redacted]	211.26	38.62	5.47	1.44	5.68	3.72	87	5.68	0 hr 0 mi	0	1.44	0	0	1864
[redacted]	193.17	26.96	7.16	0.17	7.21	0.62	0	7.21	0 hr 2 mi	0.02	0.13	0.01	0	1865
[redacted]	386.71	61.49	6.29	5.74	6.94	9.33	74	6.94	7 hr 34 m	3.82	1.92	0	0	2263
[redacted]	145.47	19.15	7.6	1.01	8.02	5.29	80	8.02	0 hr 14 m	0.17	0.83	0	0	2225
[redacted]	240.12	41.23	5.82	0.95	5.96	2.3	94	5.96	0 hr 6 mi	0.06	0.89	0.01	0	1790
[redacted]	183.93	27.59	6.67	1.55	7.06	5.63	84	7.06	1 hr 47 m	0.33	1.23	0	0	2049
[redacted]	216.09	35.45	6.09	2.21	6.49	6.22	72	6.49	1 hr 10 m	0.31	1.9	0	0	1887
[redacted]	458.12	65.48	6.92	1.31	7.06	1.99	88	7.06	0 hr 40 m	0.3	0.99	0.01	0	1864
[redacted]	0.46	0.18	2.6	0	2.61	0.7	99	2.61	0 hr 0 mi	0	0	0	0	1481
[redacted]	243.95	39.15	6.23	3.9	6.92	9.97	53	6.92	0 hr 10 m	0.09	3.49	0	0	2733
[redacted]	182.83	26.74	6.84	2.36	7.5	8.81	87	7.5	1 hr 58 m	0.77	1.58	0	0	1836
[redacted]	61.45	8.54	7.2	0.24	7.41	2.79	91	7.41	0 hr 1 mi	0.01	0.23	0	0	1854
[redacted]	642.67	87.5	7.34	0.88	7.42	1	95	7.42	0 hr 8 mi	0.05	0.81	0	1	1999
[redacted]	144.49	21.78	6.64	0.87	6.91	4.01	83	6.91	0 hr 0 mi	0	0.85	0	0	2671
[redacted]	6.46	0.77	8.39	0	8.42	0.34	98	8.42	0 hr 0 mi	0	0	0	0	2237
[redacted]	118.53	21.68	5.47	1.18	5.78	5.44	48	5.78	0 hr 35 m	0.32	0.86	0	0	4016
[redacted]	187.81	24.65	7.62	0.29	7.73	1.38	98	7.73	0 hr 16 m	0.11	0.18	0	0	1754
[redacted]	390.19	54.05	7.22	0.75	7.32	1.4	98	7.32	0 hr 6 mi	0.06	0.52	0.09	0	1791
[redacted]	517.46	87.72	5.9	5.3	6.28	6.05	77	6.28	0 hr 15 m	0.09	4.95	0.05	1	2213
[redacted]	19181.47	3059.01	6.27	90.98	6.46	2.97	84	6.46	1 day 20	18.26	66.01	1.53	6	2124

Figure A2: Lean Evaluation Criteria

Safety	Employee s have a formal avenue to openly voice, share, and regularly address safety concerns at the facility
	Safety concerns are addressed in a timely manner by a cross-functional, integrated team of employees, supervision, and management.
Safety and leadership	What level of leader is involved in the safety journey?
	What organizational levels originated, supported, and have advocated the lean implementation initiative in the facility?
Power distance	Management availability to team members. Do employees feel that management is approachable?
	What percentage of the day do Supervisors spend on the Dock, during normal working hours?
	What percentage of the day do Managers spend on the Dock, during normal working hours?
Employee recognition	Individuals who meet, exceed, or achieve objectives are recognized on a regular basis through an employee recognition program?
	Groups who meet, exceed, or achieve objectives are recognized on a regular basis through a group recognition program?
Management style	Feedback and concerns are encouraged and included before making changes and taking actions.
	Employees, Supervisors, and Managers are encouraged/empowered to try improvement ideas, using innovation and creativity to enrich job responsibilities.
	The organizational level involved in determining and leading facility, function, and CIR Goals.
Teamwork and empowerment	Daily work activities are organized into team functions.
	SME s are utilized as initial point of contact for problem-solving, resolution, and employee directing activities.
	Problem-Solving activities are organized into team functions.
	Employees are empowered, utilized, participate, initiate, and lead problem-solving activities autonomously, without significant management involvement.
Communication	There is an avenue for workers to openly share common concerns, issues, and problems regularly with other employees, supervisors, and management.
	Employee concerns and questions are addressed in a timely manner.
	Are there daily meetings with employees and supervision/management where the daily plans, performance, etc. are shared?

Figure A3: Samples from Interviews on Lean

Supervisor on how lean has affected his management style	"These guys will do anything for me, and they'll do absolutely nothing for other people. And I learned a lot of that from lean because lean has made me softer, it really has. I used to be hard as rock and now I feel like I'm a spongefiI still have that same pride but it's – my interaction with people is so much different, it's so much different. You're not treating them in a negative way or a negative manner and that's – I was hard as a rock in my numbers produced fiand if somebody didn't want to get on board with me on my team in all likelihood it probably wasn't going to be a very good day for that person. Now, it's with everybody being involved instead of just me running the show, it's totally different. Yes, are my numbers as good? Probably not, but you know what I'll take that. I firmly believe I'm a better supervisor today than what I was 6 months back."
Driver #1 on how lean has created community	"These guys now they get together, we got great relationships outside of the work environment. We've been to some of their homes. We do the activities outside of work. Even though I got Friday nights about once every month I sneak on down to Fridays and I buy them all the drink. It's just made us such a cohesive team it's incredible."
Driver #2 on how lean has increased teamwork	"I guess we haven't really been able to do too much yet – but I think the meetings and stuff have actually helped just getting people working together. So in the lean team, I think there's actually a good amount of camaraderie going on. So I think that's actually been good. Now some people I didn't really get along and stuff are working together."
Driver #1 on how lean has motivated drivers	"Since lean was introduced it was sort of like the door opening up. [Manager said] give it a chance, look at it and see what it can do. And I tell you it can produce productivity out of people that you thought would never produce. All it takes is a little bit of respect, little bit of understanding, show these guys that they're part of the operation."

Table A1: Sample construction

Sample construction	Driver-days	Sites
Total driver-days	1,137,192	275
- less early lean sites	(173,461)	(25)
- less late Q3/Q4 2013 lean sites	(130,679)	(36)
- less pre-11/25 rank posting dates	(416,593)	(72)
- less line haul routes	(76,989)	0
- less uncalibrated data	(8,781)	0
Sample	330,689	142
Sample within 5-30 window	93,913	142

Table A2: Descriptive statistics for matched sample

	Matched sample				Control	Treatment 1 (names)	Diff	Treatment 2 (IDs)	Diff
	Mean	Median	Min	Max	Mean	Mean	p-value	Mean	p-value
Site characteristics									
# sites	82	n/a	n/a	n/a	27	23	n/a	32	n/a
Lean status	0.50	n/a	0	1	0.41	0.43	0.849	0.63	0.099
Tractors / site	26.73	22.50	8	61	24.96	29.61	0.213	26.16	0.694
Distance / trip	127.79	126.34	62.08	198.69	128.26	127.66	0.945	127.47	0.918
Pre-ranking driver performance									
Miles per gallon	6.74	6.71	5.67	7.77	6.60	6.80	0.125	6.80	0.083
Gap score	2.01	1.96	0.83	5.26	2.03	2.02	0.993	1.99	0.885
Shift score	91.64	92.25	82.67	97.37	90.92	91.83	0.290	92.10	0.150
Excess idle time	0.12	0.11	0.04	0.68	0.11	0.12	0.800	0.14	0.364
Fuel lost	0.32	0.31	0.15	0.71	0.33	0.31	0.762	0.32	0.554

Table A3: Placebo test

	Actual MPG		Potential MPG	
	(1)	(2)	(3)	(4)
Post*Treatment group 1*Lean	-0.0347***	-0.0223**	0.0111	-0.0560
	(0.0102)	(0.0100)	(0.0602)	(0.0527)
Post*Treatment group 2*Lean	-0.0081	-0.0163	0.0887	0.0679
	(0.0120)	(0.0107)	(0.0595)	(0.0541)
Post*Treatment group 1	0.0065	0.0082	0.0284	0.0589*
	(0.0063)	(0.0062)	(0.0403)	(0.0339)
Post*Treatment group 2	-0.0107	-0.0024	-0.0544	-0.0353
	(0.0091)	(0.0082)	(0.0446)	(0.0348)
Post*Lean	0.0060	0.0081	-0.1129***	-0.0579
	(0.0079)	(0.0074)	(0.0399)	(0.0383)
Treatment group 1*Lean	-0.0153		-0.1943*	
	(0.0174)		(0.1018)	
Treatment group 2*Lean	-0.0204		-0.2488**	
	(0.0165)		(0.1189)	
Post	-0.0001	0.0056	-0.0779***	0.0289
	(0.0050)	(0.0059)	(0.0273)	(0.0373)
Lean	0.0220*		0.2184**	
	(0.0115)		(0.0928)	
Treatment group 1 (names)	0.0075		0.2161***	
	(0.0089)		(0.0678)	
Treatment group 2 (IDs)	0.0089		0.2222***	
	(0.0109)		(0.0819)	
Constant	-0.2591***	-0.0470***	7.0143***	6.7390***
	(0.0310)	(0.0112)	(0.1990)	(0.0264)
Date and Driver FE	N	Y	N	Date
Observations	93913	93913	93913	93913
Adjusted R-squared	0.976	0.985	0.159	0.577

Table A4: Placebo test on matched sample

	Actual MPG		Potential MPG	
	(1)	(2)	(3)	(4)
Post*Treatment group 1*Lean	-0.0341***	-0.0202**	-0.0159	-0.0272
	(0.0102)	(0.0097)	(0.0819)	(0.0695)
Post*Treatment group 2*Lean	-0.0060	-0.0050	0.1391*	0.1413**
	(0.0132)	(0.0115)	(0.0724)	(0.0608)
Post*Treatment group 1	0.0072	0.0096	0.0224	0.0595
	(0.0083)	(0.0078)	(0.0641)	(0.0499)
Post*Treatment group 2	-0.0110	-0.0081	-0.0996*	-0.0812*
	(0.0112)	(0.0105)	(0.0565)	(0.0420)
Post*Lean	0.0068	0.0035	-0.1061*	-0.0846*
	(0.0072)	(0.0061)	(0.0541)	(0.0464)
Treatment group 1*Lean	-0.0234		-0.2250*	
	(0.0242)		(0.1310)	
Treatment group 2*Lean	-0.0177		-0.2563*	
	(0.0209)		(0.1527)	
Post	-0.0027	0.0042	-0.0890**	0.0498
	(0.0058)	(0.0061)	(0.0394)	(0.0489)
Lean	0.0128		0.2627**	
	(0.0160)		(0.1145)	
Treatment group 1 (names)	0.0048		0.2984***	
	(0.0156)		(0.0890)	
Treatment group 2 (IDs)	0.0131		0.2947**	
	(0.0155)		(0.1245)	
Constant	-0.2334***	-0.0518***	6.8689***	6.6435***
	(0.0364)	(0.0136)	(0.2471)	(0.0346)
Date and Driver FE	N	Y	N	Date
Observations	60002	60002	60002	60002
Adjusted R-squared	0.976	0.985	0.159	0.571

Table A5: Instrumental variables analysis

Dependent variable:	log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.2553**	0.2962**	-4.3968**	-4.7937**	0.0837**	0.0793**	0.1175**	0.1194**
	(0.0997)	(0.1283)	(2.2285)	(2.3630)	(0.0344)	(0.0369)	(0.0502)	(0.0465)
Post*Treatment group 2*Lean	0.0261	0.0325	-0.6834	-0.9888	-0.0281	-0.0260	0.0131	0.0177
	(0.1291)	(0.1184)	(1.6693)	(1.5435)	(0.0459)	(0.0435)	(0.0403)	(0.0392)
Post*Treatment group 1	-0.0755*	-0.0703*	0.1585	0.0624	-0.0229	-0.0206	-0.0382***	-0.0302**
	(0.0441)	(0.0397)	(0.6976)	(0.6735)	(0.0145)	(0.0139)	(0.0138)	(0.0137)
Post*Treatment group 2	0.0859	0.0880	-0.7348	-0.6465	0.0451	0.0391	0.0204	0.0178
	(0.1074)	(0.0973)	(1.0684)	(0.9919)	(0.0433)	(0.0407)	(0.0318)	(0.0317)
Post*Lean	-0.0303	-0.0300	0.9783	1.0158*	0.0034	0.0033	-0.0098	-0.0101
	(0.0325)	(0.0295)	(0.6523)	(0.6034)	(0.0091)	(0.0091)	(0.0111)	(0.0101)
Treatment group 1*Lean	0.3318	0.0873	1.3078	1.9110	0.0111	-0.0518	0.0422	0.0123
	(0.2096)	(0.1675)	(2.3074)	(2.0494)	(0.0421)	(0.0497)	(0.0570)	(0.0609)
Treatment group 2*Lean	0.1667	0.2357	-2.2287	-1.2104	-0.0010	0.0309	0.0028	0.0817
	(0.1887)	(0.1840)	(2.2929)	(1.8750)	(0.0384)	(0.0349)	(0.0540)	(0.0701)
Post	0.0263	0.0191	-0.2856	-0.1774	0.0203***	0.0180***	0.0118*	0.0069
	(0.0221)	(0.0186)	(0.2523)	(0.2430)	(0.0064)	(0.0064)	(0.0065)	(0.0065)
Lean	-0.0937	-0.0900*	0.1299	0.3012	-0.0014	-0.0130	-0.0139	-0.0283
	(0.0587)	(0.0467)	(0.6914)	(0.6381)	(0.0106)	(0.0100)	(0.0159)	(0.0174)
Treatment group 1 (names)	-0.1011	-0.0597	-0.1197	-0.3691	-0.0126	-0.0135	0.0039	-0.0136
	(0.0982)	(0.0674)	(1.1465)	(0.9449)	(0.0200)	(0.0210)	(0.0257)	(0.0262)
Treatment group 2 (IDs)	-0.2553*	-0.1596	3.2883*	0.8652	0.0030	0.0029	-0.0317	-0.0540
	(0.1470)	(0.1035)	(1.7225)	(1.3275)	(0.0312)	(0.0298)	(0.0398)	(0.0387)
Constant	0.9741***	1.7826***	92.4570***	90.8150***	0.1007***	0.2884***	0.2579***	-0.0859*
	(0.0452)	(0.1297)	(0.4197)	(1.4618)	(0.0081)	(0.0371)	(0.0098)	(0.0484)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.011	0.126		0.041	0.002	0.045	0.002	0.101

Table A6: Instrumental variable analysis on matched sample

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.2880**	0.3091**	0.5184	-0.1058	0.1055***	0.1065**	0.1208**	0.1179***
	(0.1145)	(0.1201)	(1.6522)	(1.6767)	(0.0406)	(0.0414)	(0.0496)	(0.0427)
Post*Treatment group 2*Lean	-0.1089	-0.0627	2.0143	1.7536	-0.1269	-0.1009	-0.0109	-0.0108
	(0.2861)	(0.2453)	(2.1440)	(2.2801)	(0.1382)	(0.1190)	(0.0778)	(0.0798)
Post*Treatment group 1	-0.1185*	-0.0989*	-1.3615	-1.1202	-0.0399*	-0.0415*	-0.0441**	-0.0391**
	(0.0639)	(0.0553)	(1.2229)	(1.2327)	(0.0216)	(0.0214)	(0.0186)	(0.0188)
Post*Treatment group 2	0.1779	0.1703	-1.4758	-1.5799	0.1425	0.1166	0.0379	0.0406
	(0.2785)	(0.2386)	(2.0199)	(2.1887)	(0.1375)	(0.1185)	(0.0745)	(0.0766)
Post*Lean	-0.0435	-0.0424	-0.3965	-0.2954	-0.0017	-0.0042	-0.0139	-0.0125
	(0.0405)	(0.0323)	(0.4749)	(0.4401)	(0.0112)	(0.0106)	(0.0125)	(0.0127)
Treatment group 1*Lean	0.1259	0.1508	0.8134	-2.2238	-0.0030	0.0117	-0.0306	0.0677
	(0.2426)	(0.2461)	(2.2628)	(2.4720)	(0.0545)	(0.0591)	(0.0582)	(0.0892)
Treatment group 2*Lean	-0.2500	0.4302	-2.3961	-5.5245	-0.0432	0.1004	-0.0973	0.1904
	(0.4089)	(0.3118)	(3.3864)	(3.3759)	(0.0767)	(0.0801)	(0.0986)	(0.1237)
Post	-0.0425	0.0035	0.0908	0.5247	-0.0006	-0.0259	0.0039	0.0126
	(0.1235)	(0.1413)	(1.2531)	(1.4532)	(0.0303)	(0.0413)	(0.0312)	(0.0559)
Lean	0.1700	-0.2977	3.1711	3.4086	0.0405	-0.0462	0.0451	-0.1169
	(0.3858)	(0.1901)	(3.0445)	(2.1764)	(0.0730)	(0.0607)	(0.0923)	(0.0755)
Treatment group 1 (names)	0.0496*	0.0338	-0.1958	-0.0597	0.0245***	0.0214**	0.0173**	0.0109
	(0.0294)	(0.0254)	(0.2550)	(0.3086)	(0.0095)	(0.0094)	(0.0082)	(0.0089)
Treatment group 2 (IDs)	0.0135	-0.0663	-0.3308	1.0935	0.0022	-0.0217	0.0198	-0.0346
	(0.0834)	(0.0801)	(0.7739)	(0.9326)	(0.0137)	(0.0167)	(0.0187)	(0.0309)
Constant	0.8846***	1.7607***	93.3009***	89.0525***	0.0973***	0.2958***	0.2427***	-0.0746
	(0.0570)	(0.1562)	(0.4391)	(1.6999)	(0.0100)	(0.0379)	(0.0125)	(0.0553)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	60002	60002	60002	60002	60002	60002	60002	60002
Adjusted R-squared	.	0.117	.	0.046	.	0.048	0.004	0.088

Table A7: Effect of rankings on lean and non-lean sites - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.0966**	0.0963**	-1.6616	-1.4921	0.0259	0.0252	0.0465***	0.0373***
	(0.0441)	(0.0429)	(1.0370)	(1.0080)	(0.0163)	(0.0156)	(0.0156)	(0.0140)
Post*Treatment group 2*Lean	0.0263	0.0297	-1.5130	-1.4498	-0.0036	0.0001	0.0130	0.0131
	(0.0486)	(0.0464)	(0.9732)	(0.9436)	(0.0183)	(0.0167)	(0.0185)	(0.0161)
Post*Treatment group 1	-0.0414	-0.0526**	0.6129	0.6155	-0.0116	-0.0114	-0.0259***	-0.0201**
	(0.0270)	(0.0248)	(0.6369)	(0.6179)	(0.0082)	(0.0073)	(0.0093)	(0.0090)
Post*Treatment group 2	0.0289	0.0295	-0.0243	-0.0978	0.0168	0.0141	0.0065	0.0069
	(0.0354)	(0.0325)	(0.7387)	(0.7135)	(0.0127)	(0.0117)	(0.0133)	(0.0113)
Post*Lean	-0.0280	-0.0332	0.3156	0.3168	0.0066	0.0023	-0.0124	-0.0115
	(0.0328)	(0.0317)	(0.7552)	(0.7297)	(0.0121)	(0.0111)	(0.0117)	(0.0103)
Treatment group 1*Lean	0.1407	0.0377	0.3884	0.9014	0.0100	-0.0046	0.0105	0.0076
	(0.0870)	(0.0675)	(1.3413)	(1.1558)	(0.0168)	(0.0161)	(0.0227)	(0.0246)
Treatment group 2*Lean	0.0653	0.0556	-0.1648	0.8462	-0.0000	0.0080	0.0007	0.0194
	(0.0852)	(0.0692)	(1.1586)	(1.0033)	(0.0145)	(0.0157)	(0.0230)	(0.0269)
Post	0.0471**	0.0532***	1.6888***	1.6883***	0.0170***	0.0189***	0.0229***	0.0198***
	(0.0196)	(0.0167)	(0.4700)	(0.4476)	(0.0060)	(0.0053)	(0.0065)	(0.0058)
Lean	-0.0936*	-0.0628	0.7248	0.3364	-0.0034	-0.0166	-0.0121	-0.0200
	(0.0561)	(0.0487)	(0.8861)	(0.7853)	(0.0096)	(0.0107)	(0.0148)	(0.0180)
Treatment group 1 (names)	-0.0499	-0.0072	-0.0830	-0.4096	-0.0081	-0.0076	0.0038	0.0007
	(0.0553)	(0.0325)	(0.9339)	(0.7730)	(0.0101)	(0.0095)	(0.0131)	(0.0124)
Treatment group 2 (IDs)	-0.1021*	-0.0355	1.4287*	0.2472	-0.0011	0.0031	-0.0164	-0.0121
	(0.0614)	(0.0438)	(0.8018)	(0.7815)	(0.0104)	(0.0109)	(0.0142)	(0.0170)
Constant	0.9776***	1.7193***	90.3028***	87.4869***	0.0953***	0.2802***	0.2588***	-0.1257***
	(0.0428)	(0.1134)	(0.5702)	(1.3161)	(0.0071)	(0.0289)	(0.0085)	(0.0440)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	310084	310084	310084	310084	310084	310084	310084	310084
Adjusted R-squared	0.008	0.128	0.010	0.052	0.006	0.048	0.004	0.107

Table A8: Date and driver fixed effects - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.0829*	0.0822*	-1.4671	-1.4290	0.0264*	0.0228	0.0375**	0.0322**
	(0.0423)	(0.0426)	(1.0750)	(1.0707)	(0.0154)	(0.0148)	(0.0154)	(0.0144)
Post*Treatment group 2*Lean	0.0350	0.0333	-1.6876*	-1.6902*	-0.0026	-0.0013	0.0127	0.0154
	(0.0399)	(0.0402)	(1.0178)	(1.0144)	(0.0152)	(0.0148)	(0.0147)	(0.0139)
Post*Treatment group 1	-0.0516**	-0.0534**	0.7238	0.7138	-0.0095	-0.0074	-0.0251**	-0.0209**
	(0.0259)	(0.0254)	(0.6759)	(0.6738)	(0.0070)	(0.0069)	(0.0097)	(0.0097)
Post*Treatment group 2	0.0095	0.0111	-0.0248	-0.0191	0.0123	0.0107	0.0037	0.0005
	(0.0281)	(0.0276)	(0.7814)	(0.7782)	(0.0108)	(0.0105)	(0.0098)	(0.0096)
Post*Lean	-0.0310	-0.0319	0.4345	0.4371	0.0020	0.0023	-0.0128	-0.0119
	(0.0283)	(0.0288)	(0.7643)	(0.7586)	(0.0097)	(0.0093)	(0.0104)	(0.0096)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	-0.0031	-0.0047	1.0660**	1.0748**	0.0069	0.0069	-0.0033	-0.0025
	(0.0194)	(0.0187)	(0.5220)	(0.5212)	(0.0074)	(0.0074)	(0.0065)	(0.0065)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9637***	1.8205***	91.4879***	86.2097***	0.1193***	0.1670***	0.2628***	-0.0866***
	(0.0126)	(0.0411)	(0.3789)	(0.6108)	(0.0057)	(0.0105)	(0.0054)	(0.0184)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	310084	310084	310084	310084	310084	310084	310084	310084
Adjusted R-squared	0.515	0.527	0.526	0.529	0.264	0.277	0.430	0.483

Table A9: Matched analysis with date and driver fixed effects - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.0990*	0.1004*	-0.5058	-0.4726	0.0445**	0.0389**	0.0464**	0.0378*
	(0.0549)	(0.0541)	(1.4009)	(1.3983)	(0.0177)	(0.0175)	(0.0199)	(0.0194)
Post*Treatment group 2*Lean	0.0031	0.0099	-1.1112	-1.1595	-0.0131	-0.0117	0.0040	0.0025
	(0.0466)	(0.0460)	(1.4296)	(1.4241)	(0.0156)	(0.0156)	(0.0165)	(0.0166)
Post*Treatment group 1	-0.0566	-0.0582	-0.2796	-0.2884	-0.0203**	-0.0179*	-0.0290**	-0.0249*
	(0.0361)	(0.0354)	(0.7970)	(0.7936)	(0.0093)	(0.0091)	(0.0137)	(0.0135)
Post*Treatment group 2	0.0319	0.0299	-0.2026	-0.1741	0.0198	0.0174	0.0095	0.0071
	(0.0359)	(0.0353)	(1.0579)	(1.0549)	(0.0134)	(0.0134)	(0.0124)	(0.0123)
Post*Lean	-0.0375	-0.0439	0.0202	0.0464	-0.0054	-0.0041	-0.0176	-0.0126
	(0.0354)	(0.0341)	(1.1392)	(1.1336)	(0.0082)	(0.0084)	(0.0128)	(0.0130)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	0.0041	0.0048	1.5634**	1.5626**	0.0112	0.0108	0.0007	-0.0002
	(0.0245)	(0.0231)	(0.6650)	(0.6673)	(0.0097)	(0.0096)	(0.0086)	(0.0086)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9436***	1.7822***	91.8457***	86.7289***	0.1204***	0.1765***	0.2631***	-0.0862***
	(0.0167)	(0.0558)	(0.5084)	(0.7706)	(0.0074)	(0.0129)	(0.0069)	(0.0233)
Controls	N	Y	N	Y	N	Y	N	Y
Date and Driver FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	198831	198831	198831	198831	198831	198831	198831	198831
Adjusted R-squared	0.519	0.531	0.527	0.529	0.261	0.274	0.437	0.488

Table A10: Matched analysis - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1230**	0.1366**	-0.6115	-0.5885	0.0573***	0.0536***	0.0645***	0.0528***
	(0.0570)	(0.0522)	(1.4459)	(1.3966)	(0.0186)	(0.0176)	(0.0196)	(0.0187)
Post*Treatment group 2*Lean	0.0116	0.0290	-0.9394	-0.9843	-0.0025	0.0012	0.0125	0.0097
	(0.0582)	(0.0550)	(1.4392)	(1.3895)	(0.0188)	(0.0172)	(0.0227)	(0.0199)
Post*Treatment group 1	-0.0563	-0.0702**	-0.2320	-0.1899	-0.0238*	-0.0228*	-0.0333***	-0.0281**
	(0.0365)	(0.0324)	(0.7581)	(0.7241)	(0.0129)	(0.0116)	(0.0123)	(0.0118)
Post*Treatment group 2	0.0429	0.0304	-0.1244	-0.1089	0.0216	0.0176	0.0063	0.0070
	(0.0452)	(0.0429)	(1.0250)	(0.9792)	(0.0168)	(0.0156)	(0.0178)	(0.0150)
Post*Lean	-0.0524	-0.0629*	-0.2213	-0.1467	-0.0148	-0.0152	-0.0250	-0.0197
	(0.0412)	(0.0370)	(1.1712)	(1.1342)	(0.0105)	(0.0091)	(0.0151)	(0.0141)
Treatment group 1*Lean	0.0791	0.0207	0.0990	-0.6190	0.0041	0.0015	-0.0115	0.0110
	(0.1113)	(0.0900)	(1.5884)	(1.4769)	(0.0229)	(0.0221)	(0.0247)	(0.0317)
Treatment group 2*Lean	-0.0633	0.0739	0.9075	-0.0142	-0.0068	0.0076	-0.0311	0.0371
	(0.1058)	(0.0798)	(1.5747)	(1.3542)	(0.0176)	(0.0160)	(0.0265)	(0.0301)
Post	0.0626**	0.0735***	1.7184***	1.6373***	0.0265***	0.0261***	0.0308***	0.0266***
	(0.0277)	(0.0229)	(0.6430)	(0.6093)	(0.0087)	(0.0079)	(0.0093)	(0.0083)
Lean	0.0007	-0.0004	-0.3159	0.3830	-0.0024	-0.0073	0.0173	-0.0074
	(0.0786)	(0.0651)	(1.2546)	(1.1134)	(0.0124)	(0.0114)	(0.0168)	(0.0246)
Treatment group 1 (names)	-0.0248	0.0284	0.7930	0.7404	-0.0028	-0.0092	0.0014	0.0145
	(0.0701)	(0.0604)	(1.0287)	(0.9204)	(0.0149)	(0.0165)	(0.0153)	(0.0231)
Treatment group 2 (IDs)	0.0266	-0.0850	0.7728	1.0107	0.0050	-0.0025	0.0024	-0.0339
	(0.0761)	(0.0630)	(1.0768)	(1.1107)	(0.0126)	(0.0126)	(0.0181)	(0.0240)
Constant	0.8986***	1.5815***	91.0687***	88.5450***	0.0929***	0.2807***	0.2459***	-0.1567***
	(0.0562)	(0.1392)	(0.7216)	(1.5272)	(0.0086)	(0.0313)	(0.0106)	(0.0503)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	198831	198831	198831	198831	198831	198831	198831	198831
Adjusted R-squared	0.008	0.130	0.006	0.051	0.008	0.051	0.005	0.108

Table A11: Collective values sub-categories

Dependent variable:	Trust	Trust	Team	Team	Feel valued	Feel valued	Pride	Pride
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lean	0.1886	0.2315*	0.2711*	0.3193***	0.2463*	0.2873**	0.2690	0.3051**
	(0.1761)	(0.1371)	(0.1410)	(0.1141)	(0.1413)	(0.1133)	(0.1878)	(0.1455)
Constant	3.2657***	3.7042***	3.3602***	1.9307**	3.2243***	2.8677***	3.4714***	3.6234***
	(0.1733)	(0.3302)	(0.1821)	(0.9259)	(0.1648)	(0.6551)	(0.1853)	(0.5083)
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	563	563	565	565	563	563	565	565
Adjusted R-squared	0.025	0.126	0.018	0.090	0.018	0.091	0.028	0.119

Table A12: Matched analysis of collective values sub-categories

Dependent variable:	Trust	Trust	Team	Team	Feel valued	Feel valued	Pride	Pride
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lean	0.1102	0.2108	0.2269	0.3297***	0.2752*	0.3578***	0.1471	0.2439
	(0.1660)	(0.1417)	(0.1338)	(0.1102)	(0.1492)	(0.1175)	(0.1899)	(0.1661)
Constant	3.3034***	3.6955***	3.4915***	1.8946***	3.2338***	2.9563***	3.6017***	3.7583***
	(0.2261)	(0.3427)	(0.2051)	(0.6023)	(0.2033)	(0.3967)	(0.2452)	(0.3228)
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	398	398	400	400	398	398	400	400
Adjusted R-squared	0.029	0.120	0.024	0.093	0.028	0.098	0.024	0.126

Table A13: Effect of ranking on engagement

Dependent variable:	Collective	Instrumental	Trust	Team	Feel valued	Pride
	(1)	(2)	(3)	(4)	(5)	(6)
Lean*Treatment Group 1	0.0542	0.0874	0.0246	0.0074	0.0758	0.0279
	(0.2755)	(0.2458)	(0.2663)	(0.2778)	(0.3109)	(0.2696)
Lean* Treatment Group 2	-0.2591	-0.2997	-0.3220	-0.2454	-0.1832	-0.2760
	(0.4386)	(0.4661)	(0.4775)	(0.4061)	(0.4246)	(0.4708)
Lean	0.3179	0.3847*	0.3144	0.3671	0.3111	0.3509
	(0.2681)	(0.2265)	(0.2590)	(0.2798)	(0.3110)	(0.2640)
Treatment Group 1 (Names)	-0.0269	-0.1790	-0.0506	0.1598	-0.0652	0.0167
	(0.1666)	(0.1849)	(0.1725)	(0.1947)	(0.1846)	(0.1851)
Treatment Group 2 (IDs)	0.1736	0.1142	0.1860	0.2717	0.1554	0.1350
	(0.1606)	(0.1554)	(0.1659)	(0.2047)	(0.1764)	(0.1695)
Constant	3.1621***	2.7206***	3.5814***	1.6736*	2.7615***	3.5032***
	(0.4871)	(0.4608)	(0.3292)	(0.9175)	(0.6637)	(0.4921)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	561	564	563	565	563	565
Adjusted R-squared	0.126	0.079	0.127	0.089	0.089	0.115

Table A14: Matched analysis of effect of ranking on engagement

Dependent variable:	Collective	Instrumental	Trust	Team	Feel valued	Pride
	(1)	(2)	(3)	(4)	(5)	(6)
Lean*Treatment Group 1	-0.0831	-0.2282	-0.1625	0.0781	-0.0544	-0.1177
	(0.2638)	(0.2675)	(0.2737)	(0.2588)	(0.2943)	(0.2623)
Lean* Treatment Group 2	-0.4986	-0.8100*	-0.6192	-0.2021	-0.4290	-0.5379
	(0.3982)	(0.4346)	(0.4278)	(0.3782)	(0.4233)	(0.4501)
Lean	0.4296*	0.5557***	0.4421*	0.3101	0.4946*	0.4197*
	(0.2337)	(0.1918)	(0.2378)	(0.2478)	(0.2884)	(0.2175)
Treatment Group 1 (Names)	0.0963	0.1429	0.1448	0.0639	0.0151	0.1339
	(0.2310)	(0.2835)	(0.2479)	(0.2492)	(0.2394)	(0.2758)
Treatment Group 2 (IDs)	0.4260**	0.6481***	0.5258**	0.2164	0.3727	0.3867*
	(0.1835)	(0.1611)	(0.2041)	(0.2429)	(0.2206)	(0.1949)
Constant	3.2137***	2.8489***	3.5503***	1.7501**	2.9294***	3.6087***
	(0.4704)	(0.6394)	(0.4485)	(0.7324)	(0.5019)	(0.4054)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	396	399	398	400	398	400
Adjusted R-squared	0.127	0.065	0.123	0.087	0.095	0.123

Table A15: Collective values sub-categories

Category:	Trust	Trust	Team	Team	Feel valued	Feel valued	Pride	Pride
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*[Category]	0.1253***	0.1297**	0.0643*	0.0667	0.1123***	0.1079**	0.0544*	0.0621
	(0.0428)	(0.0553)	(0.0356)	(0.0408)	(0.0377)	(0.0402)	(0.0278)	(0.0378)
Post*Treatment group 2*[Category]	0.0714	0.0612	-0.0405	-0.0820**	0.0204	-0.0273	-0.0091	-0.0177
	(0.0653)	(0.0858)	(0.0392)	(0.0395)	(0.0469)	(0.0503)	(0.0385)	(0.0527)
Post*Treatment group 1	-0.4402***	-0.4783***	-0.2507**	-0.2840**	-0.3859***	-0.3958***	-0.2223**	-0.2731**
	(0.1296)	(0.1644)	(0.1086)	(0.1219)	(0.1073)	(0.1068)	(0.0908)	(0.1178)
Post*Treatment group 2	-0.1710	-0.2293	0.1780	0.2225	-0.0076	0.0556	0.0767	0.0226
	(0.2328)	(0.3298)	(0.1400)	(0.1701)	(0.1714)	(0.2082)	(0.1640)	(0.2323)
Post*[Category]	-0.0736**	-0.0717	-0.0162	-0.0086	-0.0597**	-0.0429	-0.0338	-0.0313
	(0.0356)	(0.0488)	(0.0269)	(0.0317)	(0.0290)	(0.0333)	(0.0215)	(0.0328)
Treatment group 1*[Category]	0.1063*	0.1206*	0.0776*	0.0739	0.0798	0.1001	0.1094*	0.1490***
	(0.0551)	(0.0633)	(0.0442)	(0.0522)	(0.0556)	(0.0598)	(0.0545)	(0.0527)
Treatment group 2*[Category]	-0.0832	-0.0475	0.0329	0.1253*	0.0037	0.1382*	-0.0617	-0.0103
	(0.0568)	(0.0813)	(0.0712)	(0.0644)	(0.0767)	(0.0687)	(0.0528)	(0.0594)
Post	0.2924***	0.3035**	0.1094	0.1042	0.2393***	0.2046**	0.1716**	0.1797*
	(0.1078)	(0.1422)	(0.0792)	(0.0915)	(0.0827)	(0.0876)	(0.0726)	(0.0990)
[Category]	-0.0457	-0.0620	-0.0372	-0.0372	-0.0094	-0.0330	-0.0165	-0.0354
	(0.0360)	(0.0396)	(0.0289)	(0.0296)	(0.0396)	(0.0373)	(0.0339)	(0.0368)
Treatment group 1 (Names)	-0.2910	-0.2991	-0.1903	-0.1251	-0.1936	-0.2104	-0.3190	-0.4095**
	(0.2011)	(0.2240)	(0.1757)	(0.1961)	(0.2012)	(0.2143)	(0.2129)	(0.1885)
Treatment group 2 (IDs)	0.3544*	0.2715	-0.0033	-0.2609	0.0861	-0.3069	0.2802	0.1532
	(0.1851)	(0.2691)	(0.2213)	(0.1924)	(0.2481)	(0.2246)	(0.1726)	(0.1826)
Constant	2.1385***	2.2673***	2.1134***	2.1883***	1.9967***	2.1938***	2.0225***	2.1920***
	(0.2957)	(0.3322)	(0.3076)	(0.3571)	(0.3275)	(0.3737)	(0.3069)	(0.3283)
Sample	Full	Matched	Full	Matched	Full	Matched	Full	Matched
Observations	35359	26237	35505	26383	35346	26224	35464	26342
Adjusted R-squared	0.105	0.114	0.094	0.103	0.099	0.113	0.105	0.121