Do Monetary Incentives, Feedback and Recognition matter for performance?

Evidence from a field experiment in a retail services company*

Sofia M. Lourenço

Advance Research Center,
ISEG, Technical University Lisbon,
Rua Miguel Lupi 20
1249 Lisbon
Portugal
slourenco@iseg.utl.pt

August 2010; original draft March 2010

* This paper is based on my doctoral dissertation at Harvard Business School. I am grateful for the guidance and suggestions of my dissertation committee: VG Narayanan (chair), Dennis Campbell, Srikant Datar, Jeff Polzer and Willy Shih. I also thank Jose Alvarez, Teresa Amabile, Nava Ashraf, Romana Autrey, Robin Ely, Boris Groysberg, Jan Hammond, Ranjani Krishnan, Asis Martinez-Jerez, Mike Norton, George Serafeim, Bob Simons, Bill Simpson, Tom Steenburgh, Zeynep Ton, and Ruth Wageman for their helpful comments. Seminar participants in the Accounting & Management brown bag series, Organizational Behavior Lab, and Applied Statistical Methods Seminar at Harvard Business School also provided insightful suggestions. A special thanks to Ray Fung who reviewed most of the communications with the participants and the company. I am indebted to the collaborators at the research site who made this field experiment possible. All errors remain my own.

Abstract

Given that the main paradigm in accounting is agency theory, it is not surprising that the bulk of managerial accounting research has focused on monetary incentives. However in the corporate world, monetary incentives are used in conjunction with other organizational design features such as performance feedback and recognition. But the academic literature lacks an integrated perspective on how monetary incentives, feedback, and recognition work together. Holmstrom and Milgrom (1994) have identified this need to evaluate the different incentive instruments in a holistic framework. In this paper, I answer this call for research by using a field experiment in a retail services company to analyze the performance effects of providing monetary incentives, performance feedback and recognition to sales reps. I find that the sales reps that received only monetary incentives increased their performance by 8 percentage points while the sales reps that received only recognition increased performance by 12 percentage points. These two effects are not statistically distinguishable from one another. The introduction of performance feedback did not have a statistically significant effect on performance. I find also that the group that received both monetary incentives and recognition has an increase of performance that is 11 percentage points below the sum of the performance increases of the groups that only received one of the treatments. Hence, I find evidence that monetary incentives and recognition are substitutes. This finding contributes to the literature on complementarities in the organization of the workplace (Milgrom and Roberts 1990, 1995). Finally, I find that the treatments are most effective in the bottom deciles of the performance distribution and that demographic variables – gender, age and tenure – influence the effects of the different treatment conditions.

Keywords: compensation, monetary incentives, feedback, recognition, field experiment
JEL classification: J33, L22, M41, M52

I. Introduction

In this study I conduct a field experiment to identify the performance effects of providing monetary incentives, feedback and recognition. I investigate whether these organizational design decisions are complements or substitutes, that is, whether they reinforce each other or cancel each other out. I also address the question of heterogeneity in these effects by using the level of performance and demographic variables such as gender, age and tenure.

Previous studies in the managerial accounting literature have used analytical models, archival studies, and laboratory experiments to analyze the effects of providing monetary incentives and, to a lesser extent, feedback and recognition. But the ability of these methods to identify causal effects in the corporate world has been found to be limited (Baiman 1990, Ittner and Larcker 2001, Kachelmeier and King 2002, Sprinkle 2003). In this dissertation, I use a field experiment to analyze the effects of monetary incentives, feedback, and recognition on the performance of lower-level employees. The goal is to overcome some of the limitations of the previous methods, such as endogeneity and lack of realism, by combining the rigor of an experiment with the reality of a field setting (Levitt and List 2007). Field experiments have been used for decades in organizational behavior research and more recently in economics (Levitt and List 2009). This reserch methodology allows a rigorous test of causality (Lalonde 1986, Angrist and Krueger 2001, Arceneaux et al. 2006), provides a more "natural" setting for controlled experiments and theory testing (Harrison and List 2004, Levitt and List 2007), involves real stakes for the participants as opposed to small fees for participation and performance as in laboratory experiments, and constitutes a win-win situation for academics and the collaborating research sites (Levitt and List 2009).

The theoretical motivations for this paper derive from the economics, psychology, and managerial accounting literatures. In economics, the main paradigm is the principal-agent model in which monetary incentives are used to align agents' interests with those of principals (Jensen and Meckling 1976). The argument is that monetary incentives have a disciplining or price effect because the agent will increase

effort to the extent that it is more costly to shirk (Holmstrom and Milgrom 1991). The rise of experimental and behavioral economics has led to the incorporation of other preferences in the agent's utility function (Rabin 1993, Frey 1997, Ostrom 2000, Bénabou and Tirole 2006). These other preferences may interfere with the assumed linear relationship between monetary incentives and performance (Frey 1994). Recognition, specifically in the form of awards, has only recently begun to be studied in economics despite its widespread use in the corporate world (Frey 2006). In organizational behavior research, extrinsic rewards, such as monetary incentives, feedback, and recognition, have been examined in the context of reinforcement theory (Skinner 1953, Cameron and Pierce 1994), cognitive evaluation theory (Deci 1971, 1972, Deci et al. 1999), goal-setting theory (Locke and Latham 1990), and social cognitive theory (Bandura 1999). In general, these theories predict a positive effect of extrinsic rewards on performance (Jenkins et al. 1998), but motivation crowding out and perception of control can undermine that positive effect (Cameron and Pierce 1994, Deci et al. 1999). The managerial accounting literature, which has also been part of this debate, has identified task complexity and type of incentive scheme to be moderators in the relationship between monetary incentives and performance (Bonner et al. 2000). Feedback and recognition have been studied less often in managerial accounting research. Feedback has been identified as a facilitator of learning for different incentive systems (Sprinkle 2000). The main paradigm in accounting being agency theory, it is not surprising that the bulk of managerial accounting research has focused on monetary incentives. But in the corporate world, monetary incentives are used in conjunction with other organizational design features such as performance feedback and recognition.

Overall, the literature lacks an integrated perspective on how monetary incentives, feedback, and recognition work together. The need to evaluate the different incentive instruments in a holistic framework has been previously identified (Holmstrom and Milgrom 1994). Hence, the focus of this paper is the pool of extrinsic rewards used by companies, namely, monetary incentives, feedback, and recognition, and their interrelation. My goal is to identify in a field experiment the performance effects of

providing these rewards to lower level employees, and to determine whether these extrinsic rewards reinforce each other when used jointly. I also analyze the effectiveness of the different extrinsic rewards according to the characteristics of a diverse workforce.

The research site for this field experiment is a retail services company that is hired by manufacturers to run assisted sales events in retailers' stores. These events, which are demonstrations of products to final consumers, are performed by sales reps of this third party retail services company in retailers' stores. The aim is to increase consumer understanding of the products and generate excitement around the brand and thereby boost sales in the stores.

In this field experiment, the group that received only monetary incentives increased its performance, measured as the ratio of sales relative to goals, by 8 percentage points. The group that received only recognition increased its performance by 12 percentage points, but the difference between these two estimates is not statistically significant. The group that received only feedback increased its performance by 2 percentage points, but the estimate is not statistically significant. When combined in bundles of two treatments, the effects vary between 9 and 14 percentage points and, more important, these values do not differ from the estimate for the groups that received only incentives or only recognition. Moreover, when I test whether the sum of the individual effects of incentives and recognition differs from the effect on the group that received both treatments, I find the negative difference of 11 percentage points to be statistically significant. Hence, I find evidence that monetary incentives and recognition are substitutes in this setting. This finding is corroborated by the analysis of the group that received all three treatments. The increase in the performance of the bundle group is smaller than the sum of the effects for the groups that received only one of the treatments, and the 16 percentage points difference is statistically significant.

I also find that the treatments are most effective at the bottom deciles of the performance distribution, and that demographic variables, namely gender, age, and tenure, have important interactions with the treatments, making them more or less successful in terms of their performance effects. I find that monetary incentives are effective for women but not for men, and vice versa for recognition. Older (more

experienced) reps react more to monetary incentives and performance feedback and less to recognition, and vice versa for younger (less experienced) reps. High tenure reps react positively to monetary incentives, which is not the case for low tenure reps, whereas recognition has a positive effect on the performance of low tenure reps, but does not affect the performance of high tenure reps.

This paper makes a number of contributions. First, I show that it is important to consider extrinsic rewards other than monetary incentives, which have been the exclusive component of the majority of the principal agent models used in accounting research. I find evidence that recognition can be as effective as monetary incentives without incurring in the monetary cost of a bonus. Second, I show that it is important to consider from an integrated perspective monetary incentives, feedback, and recognition in decisions related to the design of reward systems. I find evidence that monetary incentives and recognition are substitutes. Third, I show that, there being important heterogeneous treatment effects, the design of an incentive system should take into account the diversity of the workforce. I find evidence that the performance effects of monetary incentives, feedback, and recognition vary with performance level, gender, age, and tenure. Hence, the "one size fits all" rule seems not to apply to the design of incentive systems. Overall, my study contributes to the literature by providing experimental evidence from the corporate world on the importance of having an integrated perspective on extrinsic rewards and considering the characteristics of a diverse workforce when designing incentive systems.

The remainder of the paper is organized as follows. Section II presents the theoretical motivation for the treatments used in this field experiment. Section III describes the research setting, Section IV the experimental design and Section V the data. Section VI reports the results and Section VII ends with the discussion and conclusion.

II. Theoretical motivation

Monetary incentives

Standard economic theory predicts that monetary incentives will improve agent performance because the agent will increase effort to earn the bonus. The underlying mechanism is that performance is a function of effort, and that introducing desirable monetary incentives will motivate the agent to put forth greater effort, though effort is costly, to earn the bonus. This disciplining or price effect is the core of agency theory, rational choice, and utility maximization (Jensen and Meckling 1976, Holmstrom 1979, Holmstrom and Milgrom 1991). This positive relationship between monetary incentives and performance has been largely documented in workplace settings (Lazear 1986, Gibbons 1998, Lazear 1999, Prendergast 1999). Managerial accounting research has also been part of this debate (Bonner et al. 2000). I discuss their findings later in this section.

It has recently been shown by behavioral and experimental economics, however, that agents may have preferences other than money that might interfere with the assumed linear relationship between monetary incentives and performance. These other preferences have been analytically developed and empirically tested. Examples of these preferences are: fairness and reciprocity (Rabin 1993, Fehr and Gatcher 2000, Fehr et al. 2007), trust (Chen 2000, Glaeser et al. 2000, Fehr and List 2004), social norms (Ostrom 2000, Bandiera et al. 2005), and reputation and image (Bénabou and Tirole 2006, Ariely et al. 2009). Additionally, the agent may derive utility from the task itself, and thus effort may be exerted even though no monetary incentive is at stake, in which case the task is pursued for reasons of intrinsic motivation (Frey 1994). This concept of intrinsic motivation was brought from psychology. Specifically, cognitive evaluation theory argues that humans can perform an activity for its own sake, and that monetary incentives can reduce this intrinsic motivation (Deci 1971, Deci 1972, Deci et al. 1999).[1] The negative effect on performance due to a reduction on intrinsic motivation was labeled in behavioral economics as the motivation crowding out effect (Frey 1994, Frey and Jegen 2001). In psychology this

[1] In Deci's (1971) words: "One is said to be intrinsically motivated to perform an activity when he receives no apparent rewards except the activity itself." This type of motivation is the opposite of extrinsic motivation, again, in Deci's (1972) words: "Extrinsic motivation, on the other hand, refers to the performance of an activity because it leads to external rewards."

effect is known as the corruption effect (Deci 1975, 1985), hidden costs of rewards (Lepper and Greene 1978), and overjustification hypothesis (Lepper et al. 1973). But others in the psychology literature, notably the behavioral school, argue that monetary incentives have a positive effect on motivation because they provide positive reinforcement (Skinner 1953). This reinforcement will increase the frequency of the rewarded behavior and thus benefit performance. Indeed, the behavioral school argues that the negative effect of monetary incentives on intrinsic motivation is a myth (Cameron and Pierce 1994, Eisenberger and Cameron 1996, Cameron et al. 2001).

Even according to the standard principal-agent model of economics research, monetary incentives may not work. When, for example, output is unobservable or difficult to observe, monetary incentives can divert the attention and reduce the motivation of agents (Holmstrom and Milgrom 1991). Although in this case the best solution is flat wage compensation, many organizations choose to use monetary incentives to align the agents' interests with those of the principal even when such action is doomed to fail. This failure has been identified in "On the folly of rewarding for A, while hoping for B" (Kerr 1975). Hence, the properties of the measures used in incentive compensation assume an important role for contracting. Previous literature has focused on the properties of these measures as used in contracting, namely, their informativeness (Holmstrom 1979), precision and sensitivity to agents' actions (Banker and Datar 1989), and the relationship among measures (Feltham and Xie 1994).

This monetary incentives paradox has been synthesized in managerial accounting research by Bonner et al. (2000). This review of laboratory experiments finds financial incentives to have a positive effect on performance in only half of the experiments. The reasons the authors advance for this disparity are the type of task and type of incentive used. As the level of task complexity increases, the effect of monetary incentives decreases, and *ceteris paribus* the greatest likelihood of a positive effect on performance is exhibited by quota schemes, followed by piece-rate schemes, tournament schemes, and fixed rate

schemes.[2] This negative effect of task complexity on the efficacy of incentives is congruent with the finding that price effects occur in the case of simple tasks (Lazear 1999).

A field study in the retailing industry, the same setting as my field experiment, finds evidence of an association between increased store sales and the introduction of a performance-based compensation plan for sales consultants (Banker et al. 1996). Moreover, this effect persists and increases over time, although it is smaller in months with high numbers of temporary workers.

Feedback

Psychology literature, namely goal-setting theory (Locke and Latham 1990), predicts that feedback has a positive effect on performance. There are five mechanisms through which feedback can translate into greater effort and, thus, higher performance. Specifically, feedback can (1) increase goal commitment (Locke and Latham 1990), (2) serve as a self-regulatory mechanism (Bandura 1986), (3) increase the perception of self-efficacy (Bandura 1991), (4) enhance learning through its informative content, and (5) exert a positive effect on motivation (Kluger and DeNisi 1996). Both the behavioral and cognitive schools have found evidence that positive feedback increases intrinsic motivation (Cameron and Pierce 1994, Eisenberg and Cameron 1996, Deci et al. 1999) and thereby improves performance.

In managerial accounting, Luckett and Eggleton's (1991) review of the behavioral consequences of feedback argues for a contingency perspective. Feedback has to be contextualized in terms of the feedback source (psychological closeness, credibility, and power), the nature of the message (sign, frequency, and type), and individual differences (locus of control and self-esteem).

Recognition

[2] Quota schemes are characterized by a flat rate until a certain targeted level of performance, above which the agent earns a bonus. Piece-rate schemes pay a predefined amount for each unit of output. Tournament schemes link pay with the relative performance of an agent relative to that of other agents. Fixed rate schemes pay a flat rate and are not linked to performance.

Psychology theories predict a positive effect of recognition on performance, and identify several mechanisms through which this occurs. Work on human needs and motivation identifies esteem and acknowledgment as goals of human actions (Maslow 1943, McClelland 1961, Alderfer 1972), and thus employees will expend effort to obtain recognition. Recognition is also an opportunity to naturally reinforce a desired behavior (Vroom 1964, Bandura 1986), and thus the frequency of the rewarded behavior will increase. In the framework of social cognitive theory (Bandura 1999), recognition has a positive effect on performance because employees derive utility from the predictive value of recognition relative to future desired outcomes like a promotion, raise, or project assignment (Bandura 1986). The idea that rewards affect an agent's behavior by signaling the principal's private information has also been presented in economics (Bénabou and Tirole 2003). The argument is that rewards are viewed through a *looking glass self* whereby an agent makes inferences from the principal's decision, and thus two different effects are in play. On one hand, there is the direct effect in the agent's behavior due to the reward itself (e.g., a monetary incentive will have a positive disciplining or price effect). On the other hand, there is an indirect effect in the agent's behavior that is due to the agent's interpretation of the principal's decision. This interpretation can cause motivation crowding out (negative interpretation) or crowding in (positive interpretation), which will affect performance.

Recognition can thus be framed as an implicit incentive, and managerial accounting research has shown that non-financial measures not used for financial compensation are weighted in promotion and demotion decisions for lower-level managers, and that managers' behavior is consistent with these implicit incentives (Campbell 2008). Moreover, because prior literature shows that, for a given job category, significant increases in employee compensation are more attributable to promotions than to bonuses (Baker et al. 1988), recognition is important to employees as a signal of future desired outcomes.

More recently, economics research has recognized that awards, that is, non-material extrinsic compensation, have rarely been studied despite widespread use in the corporate as well as non-profit sectors (Frey 2006). Furthermore, because such awards differ substantially from monetary and other non-

monetary compensation in terms of social distinction and scarcity value, more research is needed (Frey 2007). Evidence from a vignette study shows awards to have a positive effect on willingness to contribute to a public good, that this effect increases monotonically with the value of the monetary payment or gift that comes with the award, and that this effect is lower for gifts than for payments of equal value (Neckermann and Frey 2008). The study further shows that public awards elicit greater response, and that after the award is announced willingness to contribute increases for recipients and decreases for non-recipients relative to the initial response.

Interactions between Monetary incentives, Feedback and Recognition

Although in many companies monetary incentives, feedback, and recognition coexist as organizational design choices, they have rarely been studied in an integrated manner. The bulk of research in economics is on monetary incentives in the framework of agency theory. In psychology, different streams of research study monetary incentives, performance feedback, and recognition, but rarely take a comprehensive approach. Examples of streams of research on these topics are expectancy theory, reinforcement theory, goal setting theory, and cognitive evaluation theory. Additionally, research in psychology focuses neither on the performance effects of these organizational design choices nor on comparing them. Hence, the effect of combining monetary incentives, performance feedback, and recognition (e.g., do two or more features that are combined reinforce each other or cancel each other out) has not been studied despite its widespread use. In the wording of economics, the question is whether monetary incentives, performance feedback, and recognition are complements or substitutes in terms of their performance effects. I contribute to the literature on complementarities in the organization of the workplace by addressing this question in a field experiment. I use the Milgrom and Roberts' (1995) definition: "The notion of complementarity we use is due to Edgeworth: activities are *Edgeworth complements* if doing (more of) any one of them increases the returns to doing (more of) the others." Previous research on this topic shows analytically that technology, structure, strategy, job design, and incentive systems can be complements (Milgrom and Roberts 1990, 1995; Holmstrom and Milgrom

1994), and that objective and subjective measures can be complements in explicit and implicit incentive contracts (Baker et al. 1994). This latter theoretical model is supported by empirical research that shows objective group piece rates (explicit contracts) and subjective bonus plans (implicit contracts) to be more effective when implemented together (Ichniowski et al. 1997). More generally, empirical findings from archival data studies show job design, human resource management practices, and compensation schemes to have important complementarities (Ichniowski et al. 1997; MacLeod and Parent 1999). But these studies do not explain "why only some firms implement innovative workplaces with higher effort discretion, as this enhanced productivity in the cases that were studied" (Bartling et al. 2010). Complementarities among high effort discretion, rent-sharing, screening opportunities, reputation formation, and competition are shown via a laboratory experiment to be driving forces behind "high-performance work systems" (Bartling et al. 2010).

Overall these studies show the importance of analyzing different components jointly in job design decisions. Observe Holmstrom and Milgrom (1994):

> We have also focused on a limited set of incentive instruments, and one may well ask whether other instruments, such as promotions, implicit incentives, or simple verbal encouragement, are not equally important. They probably are, and they would have to be included in any purportedly complete study. As new instruments are added, however, the key according to our theory is to evaluate them not in isolation, but as part of a coherent incentive system.

I respond to this call by analyzing jointly the performance effects of introducing monetary incentives, feedback, and recognition. The majority of studies in managerial accounting research also analyze a single variable. One of the exceptions is a laboratory experiment that used both performance feedback and monetary incentives (Hannan et al. 2008). The results show relative performance feedback to increase the mean performance of participants compensated under an individual incentive scheme, but decrease the

mean performance of participants compensated under a tournament incentive scheme if the feedback is more precise. This feedback manipulation is different from the intervention used in my field experiment, as my feedback treatment compares individual performance to goals rather than among peers. Another experiment, aimed at identifying the effect of incentive contracts on learning and performance, shows incentive-based compensation to improve performance on a multi-period cognitive task only after considerable feedback and experience (Sprinkle 2000). Studies on goal-setting theory also link improvements in performance to the combined use of monetary incentives and feedback (Latham and Locke 1991).

III. Research Setting

The research site of my field experiment is a U.S. retail services company with two divisions, Merchandising and Assisted Sales, that have different functions. Employees of the Merchandising division, which is hired by manufacturers to do product replacement and product display in retail stores, do not interact with consumers. Employees of the Assisted Sales division, which is hired by manufacturers to conduct product demonstrations in retail stores, interact with consumers. The rationale for manufacturers paying a third party company to conduct product demos is that these events boost sales by helping customers understand a product and generating excitement around the brand. The company's employees can work full- or part-time and, because of the different tasks involved, generally work exclusively for one division or the other. Sales reps employed by the Assisted Sales division work throughout the country and interact with the company primarily through field managers responsible for specific regions. The study participants of the field experiment were exclusively part-time sales reps employed by Assisted Sales.

The fact that sales reps do not interact with one another, and there is thus no contamination of the treatment groups, makes the company an ideal site for this field experiment. The only mechanism that allows the reps to interact with each other is a Yahoo Group. This forum allows the reps to share experiences, report issues, and ask product questions. The management team monitored carefully this website to assure that no contamination of the treatment groups occurred, and no comments about the intervention were posted. The post-experimental questionnaire confirmed that no contamination had occurred since no rep was aware of other treatment conditions. Hence, the experiment does not address spillover effects or peer pressure, but it entirely eliminates concerns about contamination among the treatments that may bias the results.

The company is an example of an organization with weak controls for low level employees. Before the experiment, sales performance was not computed for individual reps. The company focused instead exclusively on store performance, and the main concerns of the management team were scheduling and reporting. The reps are required to conduct demonstrations on dates set by the manufacturer to coincide with important retail dates, and to provide through online reporting extensive details of each event (e.g., number of interactions with customers, number of products bought by customers while the sales rep was at the store, descriptions of issues related to the product raised by customers, problems with products in stock, product display, and prices).[3] The reps are also encouraged to upload photos of product displays. The execution of the schedule can be controlled by the store managers of the retailer and reps' reports are used by the company's management team to provide weekly updates to manufacturers. This real time field information can be used to resolve issues with product stock, prices, and displays, to incorporate consumer feedback in future product development, and to inform future marketing strategies by comparison with "hard" store sales data. Although reps and field managers were thus not focused on store sales, top management went to great lengths to document to manufacturers a return on their investment in

[3] An event is a four-hour period of assisted sales for a specific product at a particular retail store.

Assisted Sales. This analysis was usually done by comparing the performance of covered and uncovered stores.

This site is also attractive from a research standpoint because the baseline situation is the absence of monetary incentives, feedback, and recognition. Sales reps earned $17 per hour, stores had no assigned goals, no bonuses were paid, and no feedback on store sales performance was provided, nor was any type of recognition. The reps' interactions with the field managers were concerned mainly with event assignments, scheduling issues, reporting, and training.

Additionally, reporting in the retailing industry is done weekly, not monthly as in many other industries, which allows more frequent communication with participants. This is an important advantage of the research site.

IV. Experimental design

The field experiment is a between subjects 2 x 2 x 2 factorial design, and the variables are the presence or absence of monetary incentives, performance feedback, and recognition. There are thus three groups with only one of the treatments, three pairwise combinations of the treatments, and one full combination of the three treatments. The experimental design is detailed in Figures 1 and 2.

At the beginning of the experiment, field managers informed reps via e-mail of the new programs the company was implementing. This e-mail differed from rep to rep according to the random assignment to a treatment (Appendix 1). The reps were unaware that other programs, different from the one described in their e-mail, were being implemented simultaneously. Also communicated in the e-mails were sales goals defined by the management team for each store. These goals were based on past performance and a seasonality factor. Three levels of goals, expressed in terms of units of products to sell, were specified: "Goal," "Stretched Goal," and "Super Stretched Goal." The control group also received an e-mail with

store goals. All reps, except the ones in the control group, were informed weekly via e-mail of their performance according to their treatment condition.

Monetary Incentives

Reps in the incentive condition could earn a weekly bonus. The bonus was conditional on meeting sales goals taking into consideration the average performance of the events conducted in that week. The total amount of the weekly bonuses was paid at the end of the campaign. An extra $10 (for meeting the Goal), $20 (for meeting the Stretched Goal) or $30 (for meeting the Super Stretched Goal) could be earned each week. These values could be as little as 4% or as much as 44% of the weekly flat compensation, the difference reflecting the number of hours worked and type of bonus earned (Goal, Stretched Goal, or Super Stretched Goal). The reps were informed weekly by the payroll department, via e-mail, of any bonuses earned in the previous week (Appendix 2). The week delay is due to the retailer's report system.

Performance feedback

Reps in the feedback condition received information about sales and goals, by product category, for the stores worked in the previous week (Appendix 3). This information was sent weekly by the operations department via e-mail. The same week delay applies as described in the monetary incentives condition.

Recognition

Reps in the recognition condition participated in a special acknowledgement program that awarded certificates, bearing the rep's name and the award category, signed by the CEO.[4] The acknowledgement program was complemented with a public announcement in the corporate newsletter, again with the rep's

[4] Although the design of this condition and the initial e-mail sent to reps indicated that the certificates were to be signed by field managers, in the company's implementation, the handwritten signature was that of the CEO and not of a field manager.

name and the award category. The certificate was delivered to the reps and the announcement was made only after the end of the experiment. The acknowledgement program was designed to mimic the incentives condition. Hence, a rep who met the "Super Stretched Goal" was classified as a Gold Rep, a rep who met the "Stretched Goal" a Silver Rep, a rep who met the "Goal" a Bronze Rep, and a rep who did not meet the "Goal" a Promising Rep. Reps received weekly from their field managers an e-mail informing them of their award category for the previous week (Appendix 4). The same week delay described for the other conditions applies.

V. Data

Reps were randomly assigned to the treatment groups at the start of the experiment. Pre-experimental data was available for some but not for other reps, and some reps assigned to conditions ended up not working on the manufacturer campaign that was the experiment scenario. These two situations imply that the groups are not precisely the same in the pre-experimental and experimental periods. But due to their random assignment, reps' expected performance and characteristics should not differ among treatment groups. A group of reps not on the initial list worked on the campaign that was the experiment scenario but was dropped from all analyses because they were not randomized. These reps are either recently hired employees of Assisted Sales or Merchandising employees who conducted a couple of events for Assisted Sales. These Reps are identified as group X. The distribution of reps by group is shown in Table 1, Panel A.

The groups are similar in terms of gender, age and tenure as presented in Table 2. In the bottom of the table a Chi2 test for the equality of proportions and an F test for the equality of the means across the groups do not reject the null hypothesis.

The field experiment was run for a two-month campaign for two consumer electronics products, product A and product B. The price range of product A is 10 times that of product B. This campaign focused on two retail chains but only the larger had data available before the start of the experiment,

enabling store goals to be set. Hence, the experiment was run exclusively for events at stores of that retail chain, and this fact was made clear in all communications with participants. Due to a problem with retailer-manufacturer reporting, reps were not provided with information about bonuses, store sales, and award categories during the first four weeks of the experiment. Before reporting on a weekly basis commenced for the latter four weeks of the experiment, the reps were reminded, via an e-mail, of their treatment condition and store sales goals. A detailed timeline of the field experiment is presented in Figure 3.

The outcome variable is weekly rep performance. Because reps can work multiple events in a given week, the performance variable was computed as the average of the ratios of sales to goals per event conducted within the week. This ratio eliminates any effects of store size. I considered in this average only events for product B, as product A had no events during the last four weeks of the experiment, the ones with reporting. Additionally, the two products had different time periods (some weeks had only events for product A, some weeks only events for product B) and success rates (goal or above was achieved by the reps in 40% of the events for product A but only in 7% for product B). Combining the two products under these circumstances would have occasioned unnecessary noise in the outcome variable. The number of reps with data for product B is shown in Table 1, Panel B. The conclusions from Table 2 do not change when the sample is restricted to reps with data available for product B.

The outcome variable is still noisy because sales data and goals are for a span of a week, during which a rep can work as little as four hours (one event) or as many as twenty-four hours (six events). Moreover, these events can be in the same store or in different stores, and rep-store allocation can vary week to week, and different reps (with different treatments or not) can work in the same store during the same week (though only 51 of 2,859 total events for the combination manufacturer/retailer/product that was the object of the experiment occurred in the same store-week with different reps). These problems, because they do not differ by treatment group, will not drive the results. If anything, because they add noise to the outcome variable, they will bias against finding statistically significant treatment effects.

VI. Empirical Tests and Results

Table 3 describes the outcome variable for all the weeks of the data available, as well as for the pre-experimental and the experimental periods. Given the industry of the research site, retailing, and the campaign product, consumer electronics, there are strong week effects. Some weeks include important dates for retailing, e.g. Easter (weeks 14-15), Mothers' Day (weeks 18-19), and Fathers' Day (weeks 24-25). The number of reps working per week also varies with the retailing calendar, as reflected in the distribution of observations per week. Figure 4 illustrates this variation.

Table 4 presents the weekly mean performance of the groups, as well as the performance through the pre-experimental and the experimental periods. The mean performance of each group per week is depicted in Figure 5. I constrained the data to weeks in which all groups had reps working. The groups' performance, as can be seen in the figure, was similar until the beginning of the experiment (week 19), and hence it is possible to identify weekly trends. The groups' performance during the experimental period is more heterogeneous, and thus it constitutes preliminary evidence of treatment effects.

To identify the treatment effects I used the following model:

$$Average\ Sales/Goals_{it} = \alpha_0 + \sum_{k=1}^{k=7} \beta_k\ Treatment\ Groups_{ik} + \sum_{k=1}^{k=7} \gamma_k\ Treatment\ Groups_{ik} * Study\ period + \sum_{t=1}^{t=12} \omega_t\ Weeks_t + \varepsilon_{it}$$

(Equation 1)

where i represents sales reps, t represents weeks, and k represents treatment groups. $Average\ Sales/Goals_{it}$ is the performance of rep i in week t, $Treatment\ Groups_{ik}$ are dummy variables for each treatment group from B to H (group A, the control group, is omitted) according to the random assignment of rep i, $Study\ period$ is a dummy variable that is equal to 1 if week>18 and 0 otherwise, $Weeks_t$ are dummy variables for each week of data (the first week is omitted and one week for which there are no observations is dropped), and ε_{it} is the error term for rep i in week t. The dummy variable for the study period (with no interaction with the treatment groups) is omitted from the regression because it is perfectly collinear with the dummy variables for weeks in the experimental period.

This specification controls for week effects and ex-ante group differences. The model computes the treatment effects by calculating the differences between the experimental and pre-experimental periods. The control group is omitted from the regression, but allows the identification of the week effects. This specification also allows a comparison of the different treatment effects by using F tests. Table 5 presents the results of this model computed using OLS pooled regression with clustered standard errors by rep.

Equation 1 in column 1 is estimated with all the available data. The column shows a positive effect on performance (sales relative to goals) for all treatment groups, and no statistically significant differences between the treatment groups and the control group in the pre-experimental period. The performance increases for the groups that received only one treatment are 8.2 percentage points (pp) for Group B (incentives), 2.2 pp for Group C (feedback), and 11.6 pp for Group D (recognition). For the bundle groups that received combinations of treatments, the performance increases are 10.5 pp for Group E (incentives and feedback), 8.9 pp for Group F (incentives and recognition), 13.7 pp for Group G (feedback and recognition), and 6 pp for Group H (incentives, feedback, and recognition). These values are economically significant for a product with a price range of $100-$500 that sold 74.824 units during the experimental period in the retail chain of the experiment. Moreover, in the pre-experimental period sales were only 40% of the goals, and hence the overall treatment effect of 8.3 *percentage points*, when all treatment groups are pooled and compared to the control group, represents a 20 *percent* increase in

performance. The treatment effects are statistically significant for all groups except C (feedback) and H (incentives, feedback, and recognition); though the coefficient of Group H is only marginally not significant (the p value of the two-sided t-stats is .11). F tests show that the coefficients of the treatment groups are not statistically different from one another in the experimental period, except for the comparisons between Group C and groups D, E, F, and G, and between groups G and H. I also use F tests to compare the sum of the effects of the groups that received only one treatment with the effects of the bundle groups that combine two or more treatments. The coefficients of the bundle groups receiving feedback and one of the other treatments are statistically no different from the sum of the coefficients of the individual treatment groups. Economically, the differences are quite small (-0.1 pp for the combination of incentives and feedback, +0.1 pp for the combination of feedback and recognition). These results are not compelling evidence of the independence of feedback relative to incentives and recognition because feedback by itself is not statistically significant. The coefficient for the combination of incentives and recognition (8.9 pp) is 10.9 pp smaller than the sum of the individual coefficients of the groups that received only incentives (8.2 pp) or recognition (11.6 pp). This difference is significant at the 10% level (F-stats=3.7). This result provides evidence that incentives and recognition are substitutes; the effect of one crowds out the effect of the other when they are used jointly. The coefficient of Group H, the bundle group that combined the three treatments, is 16 pp smaller than the sum of the individual coefficients of the groups that received only incentives (8.2%), feedback (2.2%), or recognition (11.6%). This difference is significant at the 5% level (F-stats=4.13). This comparison, however, ignores interactions between any two of the three treatments. To overcome this limitation, in untabulated results I compare Group H with the sums of groups E and D, F and C, and G and B. These tests reveal a negative and statistically significant effect of adding recognition to a combination of incentives and feedback. Much the same occurs when incentives are added to a combination of feedback and recognition. There is, however, no statistically significant difference when feedback is added to a combination of incentives and recognition. These results are driven by the previous negative interaction between recognition and incentives and the nonsignificant effect of feedback by itself.

Because the treatments were introduced in the middle of week 19 and sales data are reported for the entire week, I exclude this week from the analysis and again estimate equation 1. The results are reported in column 2 of Table 5. Save for Group G, the coefficient estimates are larger in this specification than in the previous model. The main difference is that the coefficient for Group H is now statistically significant (9.6 pp, significant at the 5% level). The coefficient for Group C (6.1pp), although still not statistically significant in a two-sided test, is weakly significant in a one-sided test (significant at the 10% level). In this specification the F tests for the pairwise comparisons of the coefficients are never statistically significant. Results of the F tests that compare the effects of the bundle groups with the sum of the effects of the groups that received only one treatment are similar to those reported in column 1: negative interaction for incentives and recognition (but now statistically significant at the 5% level) and nonsignificant interactions with feedback.

How the non-reporting period influenced the experiment is also a concern, and I address it by estimating equation 1 excluding the reporting period (column 3) and excluding the non-reporting period (column 4). From these columns, I conclude that the treatment effects reported on columns 1 and 2 come mainly from the reporting period. In the non-reporting period the treatment effects are all not statistically significant, except for groups D and G, whereas in the reporting period all coefficients are statistically significant, except for Group C. As in the previous models, this comparison is made against the pre-experimental period. To address the question of whether reporting added an effect relative to non-reporting, I added to equation 1 an interaction of the treatment groups in the experimental period with the reporting period, a dummy variable equal to one if week>22 and zero otherwise. In untabulated results, I find none of the interactions (treatment group * study period * reporting period) to be significant. But when I use F tests to compare the coefficients of each treatment group in the pre-experimental period with the sum of the coefficients in the experimental period (study period and reporting period) for the same group, all comparisons save for Group C (feedback) are significant. This result is consistent with the conclusions reported in columns 1 and 2. The results of columns 3 and 4 show that reporting allowed for

stronger treatment effects than just the inducement introduced at the beginning of the experiment. It remains an empirical question, however, whether these effects are due to reporting or to the time period in the experiment (four weeks had already passed). Because my original experimental design did not have a non-reporting period, the data does not permit the disentangling of these alternative explanations.

Robustness checks

The analysis of the three treatment effects and their interactions can be done directly via a regression in which the dummies for the groups are replaced by dummies for the three treatments and their interactions. In untabulated results, I reach the same conclusions: incentives and recognition have a positive and statistically significant treatment effect, and the only statistically significant interaction is between incentives and recognition, with an estimated value of -10.9 pp significant at the 10% level. This regression with dummy coding of the treatments is different from a regression with effect coding of the treatments because the latter mimics the ANOVA intuition. An ANOVA compares the groups that receive a treatment with those that do not. For example, in the case of incentives the groups that receive incentives (groups B, E, F, and H) are compared with those that do not (groups A, C, D and G). This is different from the dummy coding in which every treatment group is compared with the control group (Group A). There being in this field experiment for all treatment groups a positive effect on performance relative to the control group, the effect coding regression will have weaker effects (four treatment groups will be compared with three treatment groups plus the control group). Effect coding implies that observations subject to a given treatment are assigned a value of +1 and observations not subject to the treatment a value of -1. Because the range of variables with effect coding is two, the size of the coefficients is half of those estimated with dummy coding, in which the range is one (the variables are coded 0 or 1). In untabulated results, I find that with effect coding only recognition has a statistically significant treatment effect in the form of a 2.4 pp increase. The effect of monetary incentives is a 0.8 pp

increase that is not statistically significant. The interaction between monetary incentives and recognition has an estimated coefficient of -3.4 pp that is statistically significant at the 1% level. Again, this negative interaction is evidence of a substitution effect between monetary incentives and recognition.

To address the concern that the groups are not exactly the same in the pre-experimental and experimental periods, I constrain the sample to the 261 reps with data available for both periods. Making the groups exactly the same pre- and post-experiment eliminates any possibility that the results are driven by a difference in the performance of reps who worked only during the experiment and those who worked only during the pre-experimental period. This should not be a concern in any case because of the random assignment of reps to the experimental groups. In untabulated results, I obtain results similar to those obtained in column 1 of Table 5: all treatment effects estimates are positive and statistically significant, except for groups C and H, and the coefficients vary between 7 and 14 pp.

To address the concern that pre-treatment levels of performance might be driving the results, although they are not statistically significant I exclude observations prior to the experiment. This specification also eliminates any concern about trends in the data not accounted for by the week fixed effects, although any learning should be identical across groups inasmuch as they do not differ in terms of age (proxy for working experience) and tenure in the company. In this specification, treatment effects reflect a direct comparison with the control group and not pre-post change within each group as in Table 5. The results of this specification, estimated with 704 observations of 304 reps, are similar to the results reported in column 1 of Table 5: all treatment effects estimates are positive and statistically significant, except for Group C, and the coefficients vary between 8 and 11 pp. The only difference relative to column 1 of Table 5 is that Group H has an estimated coefficient of 8 pp that is significant at the 5% level.

Finally, to address the concern that differences among reps might be driving the results, notwithstanding the random assignment of reps to the experimental groups and clustering of standard errors by rep in all specifications, I use a specification with fixed effects by rep and a specification with

random effects by rep. My findings are robust to these specifications. In the fixed effects specification, except for groups C and H, all treatment groups have a positive effect on performance and the statistically significant estimates vary between 7 and 15 pp. In the random effects specification, the results are identical to the fixed effects, save that Group H now has an estimate of 6.8 pp statistically significant at the 5% level.

Quantile regression

To analyze the distribution of the treatment effects, I ran quantile regressions for each decile of the outcome variable. The estimated coefficients for each regression are the treatment effects based on a comparison of the same decile for each group. Table 6 has the estimates of these quantile regressions with the treatment effects at every decile. In this specification I use bootstrap standard errors per rep with 100 repetitions. The quantile regressions use the 1,509 observations of the 359 reps as in column 1 of Table 5. From this table, I conclude that the treatment effects come mainly from the bottom of the distribution, the coefficients being larger and more often statistically significant in deciles 1 to 4 than in deciles 6 to 9. Hence, the treatments are more effective at low performance levels.

Heterogeneous treatment effects

Whether there is heterogeneity in the treatment effects is also important. Treatments may have a uniform effect on reps with different individual characteristics or be more or less successful with reps who share a specific characteristic. In this section, I analyze heterogeneous treatment effects based on reps' demographic characteristics, gender, tenure, and age, collected by the research site's HR department. Table 7 reports descriptive statistics for these demographic variables.

I identify potential heterogeneous treatment effects by interacting the demographic variables with the treatments using four different specifications. The first (equation 2) considers all the treatments together to assess whether the effect of providing any kind of treatment interacts with reps' demographic characteristics. This specification has the advantage of power because all the treatment groups are together, but ignores any differences among treatments. Moreover, if treatments interact with demographic variables in significant but opposite ways (one positive and one negative), the interaction in this specification may not be significant.

$$Average\ Sales/Goals_{\ it} = \alpha_0 + \beta_1\, Demographic_i + \beta_2\, Demographic_i * study\ period + \beta_3\, Treatment_i + \beta_4\, Treatment_i * Demographic_i + \sum_{t=1}^{t=12} \omega_t\, Weeks_t + \varepsilon_{it}$$

(Equation 2)

where $Demographic_i$ is the demographic characteristic of rep i (gender, age, or tenure) and the other variables are the same as defined in the previous specification.

The second specification (equation 3) considers each treatment group individually. This specification has the advantage of using all the information available, but has lower power. Additionally, interpreting some of the interactions between the demographic variable and the groups that combine two or more treatments is difficult.

$$Average\ Sales/Goals_{it}$$

$$= \alpha_0 + + \beta_1\ Demographic_i + \beta_2\ Demographic_i * study\ period$$

$$+ \sum_{k=1}^{k=7} \delta_k\ Treatment\ Groups_{ik} + \sum_{k=1}^{k=7} \eta_k\ Treatment\ Groups_{ik} * Demographic_i$$

$$+ \sum_{k=1}^{k=7} \gamma_k\ Treatment\ Groups_{ik} * Study\ period + \sum_{k=1}^{k=7} \phi_k\ Treatment\ Groups_{ik}$$

$$* Study\ period\ * Demographic_i + \sum_{t=1}^{t=12} \omega_t\ Weeks_t + \varepsilon_{it}$$

(Equation 3)

where $Treatment\ Groups_{ik}$ are dummy variables for each treatment group from B to H (Group A, the control group, is omitted) according to the random assignment of rep *i*, and the other variables are the same as defined in the previous models.

The third specification (equation 4) uses only the three treatments, monetary incentives, feedback, and recognition, and ignores interactions among them. This specification has the advantage of keeping the three treatments, which equation 2 does not, and of sample size and simplicity compared to equation 3. The disadvantage is that interactions among the treatments are ignored. I do not consider the interactions among the treatments in this dummy coding regression to avoid the problem of multicollinearity.

$$
\begin{aligned}
&Average\ Sales/Goals_{it} \\
&\quad = \alpha_0 + \ + \beta_1\, Demographic_i + \beta_2\, Demographic_i * study\ period \\
&\quad + \delta_1\, Monetary\ Incentives_i \ + \delta_2\, Feedback_i + \delta_3\, Recognition_i \\
&\quad + \eta_1 Monetary\ Incentives_i * Demographic_i + \ \eta_2\, Feedback_i * Demographic_i \\
&\quad + \ \eta_3\, Recognition_i * Demographic_i + \ \gamma_1 Monetary\ Incentives_i * Study\ period \\
&\quad + \ \gamma_2 Feedback_i * Study\ period \ + \ \gamma_3 Recognition_i * Study\ period \\
&\quad + \phi_1\, Monetary\ Incentives_i * Study\ period \ * Demographic_i + \phi_2\, Feedback_i \\
&\quad * Study\ period \ * Demographic_i \ + \phi_3\, Recognition_i * Study\ period \\
&\quad * Demographic_i + \sum_{t=1}^{t=12} \omega_t\, Weeks_t + \ \varepsilon_{it}
\end{aligned}
$$

(Equation 4)

where $Monetary\ Incentives_i$ is a dummy variable that is equal to 1 if the rep received monetary incentives as a treatment (i.e., was in group B, E, F, or H) and 0 otherwise, $Feedback_i$ is a dummy variable that is equal to 1 if the rep received feedback as a treatment (i.e., was in group C, E, G, or H) and 0 otherwise, $Recognition_i$ is a dummy variable that is equal to 1 if the rep received recognition as a treatment (i.e., was in group D, F, G, or H) and 0 otherwise, and the other variables are the same as defined in the previous models.

The fourth specification (equation 5) employs a regression with effect coding. This specification supports the analysis of the main treatment effects and their interactions without incurring the problem of multicollinearity.

$$Average\ Sales/Goals_{it}$$

$$= \alpha_0 + \ + \beta_1\ Demographic_i + \beta_2\ Demographic_i * study\ period$$

$$+ \sum_{k=1}^{k=7} \delta_k\ Treatments\ \&\ Interactions_{ik} + \sum_{k=1}^{k=7} \eta_k\ Treatments\ \&\ Interactions_{ik}$$

$$* Demographic_i + \sum_{k=1}^{k=7} \gamma_k\ Treatments\ \&\ Interactions_{ik} * Study\ period$$

$$+ \sum_{k=1}^{k=7} \phi_k\ Treatments\ \&\ Interactions_{ik} * Study\ period\ * Demographic_i$$

$$+ \sum_{t=1}^{t=12} \omega_t\ Weeks_t + \ \varepsilon_{it}$$

(Equation 5)

where $Treatments\ \&\ Interactions_{ik}$ are the variables for each of the three treatments (monetary incentives, feedback, and recognition) and the interactions among them, three two-way interactions and one three-way interaction. I use effect coding for these variables, assigning a value of +1 if the observation received the treatment and a value of -1 if the observation did not receive the treatment. An effect coding regression supports the comparison of treatment groups that received a particular condition with groups that did not receive that condition but may have received some other treatment condition. The other variables are the same as defined in previous models.

As in equation 1, all of these specifications control for week effects and ex-ante group differences. The model computes the treatment effects by calculating the difference between the experimental and pre-experimental periods relative to the control group that is omitted from the regression, but allows identification of the week effects.

Table 8 presents the results for gender. The variable used is a dummy coded 1 when the gender is female and 0 for male. Column 2, which estimates equation 2, shows the interaction between the overall treatment and the female dummy variable to be positive (10 pp), but not statistically significant. In the analysis of the groups (column 3, which estimates equation 3), the only significant interaction is that of the female variable with Group F (incentives and recognition). The main effects regression (column 4, which estimates equation 4) and regression with effect coding (column 5, which estimates equation 5) show monetary incentives to be the only treatment for which interaction with the female variable is statistically significant. The estimate of this interaction in column 4 is 8.1 pp, significant at the 10% level, and in column 5, 4.6 pp, significant at the 5% level. Due to the range of two in the effect coding regression, estimates of the coefficients of the treatments are half those derived from a dummy coding regression, and hence the two columns' estimates of the interaction are similar. In both these specifications, the F test for the monetary incentives condition for female is statistically significant, whereas the T test for the monetary incentives condition for male is not. Monetary incentives thus have a positive effect on performance for women but not for men. The opposite result is obtained for the recognition condition. The F test for the recognition condition for female is not statistically significant, whereas the T test for the recognition condition for male is. Recognition thus has a positive effect on performance for men but not for women. These puzzling results warrant further investigation in future research focused on gender differences.

The demographic variable age is analyzed next on Table 9. Age is a proxy for working experience (correlation of .87, significance <.001) and, to a lesser extent, for experience in the retailing industry (correlation of .5, significance <.001). Working experience and experience in the retailing industry are self-reported measures that were collected via the pre-experimental questionnaire. The pre-experimental questionnaire was completed by 272 reps out of 437; hence the response rate is 62%. I standardize the age variable so that the main effects of the experimental treatments (no interaction with the age variable) are

meaningful.[5] Specifically, the effects of the experimental treatments occur at the mean value of the age variable, and the interactions represent the added effect of one standard deviation above or below the mean of the age variable. Table 9 shows a positive and statistically significant interaction with the treatments, 7.2 pp, significant at the 1% level (column 2). When the interventions are separated into the different treatment groups (column 3), the significant interactions are with the bundle groups (E, F, and G). The main effects treatment analysis shows this to be due to the monetary incentives and feedback conditions (column 4), 4.1 pp and 5.3 pp, respectively. The effect coding regression (column 5) confirms these results, albeit weakly for the monetary incentives condition. Overall, the table shows the reaction to monetary incentives and performance feedback to be greater for older (more experienced) reps (one standard deviation above the mean) than for younger (less experienced) reps. The recognition condition, which is statistically significant by itself, has no statistically significant interaction with age even though the coefficient estimates are negative. The F stats for the sum of the treatment condition and the treatment condition interacted with age are statistically significant for both the monetary incentives and feedback conditions, which is not the case with the F stats for the sum of recognition and recognition interacted with age, significant at the 10% level in a two-sided test in model 4 and not significant in model 5. A possible explanation for these results is that older (more experienced) reps need less positive reinforcement from their managers concerning the job they are doing (recognition condition), and thus exhibit less reaction to this treatment than younger (less experienced) reps.

Table 10 presents the results for the tenure variable. Because reps work part-time and the Assisted Sales division is relatively new in the company, tenure has a low correlation with age (.14, significance <.05). I standardized the tenure variable for the reasons presented above for age. The interaction between the overall intervention and tenure is not statistically significant (column 2), but when I separate the intervention into the different treatment groups I identify statistically significant (and negative) interactions for groups D and G. The main treatments analysis (column 4) shows a positive interaction

[5] With no standardization, the treatment effects occur when the age variable is zero, which is never the case. Hence, the estimation would be out of the sample.

between monetary incentives and tenure, which is corroborated by the effect coding regression (column 5), which also shows a negative and statistically significant interaction between recognition and tenure. Overall, the table shows that reps with tenure above the mean respond positively to monetary incentives while reps with low tenure do not respond to this treatment. The F test for the sum of the effect of monetary incentives and the interaction between monetary incentives and tenure is positive and statistically significant; while the T test for the monetary incentives treatment is not statistically significant. The table further shows that reps with longer tenure do not respond to recognition while reps with low tenure do. The F test for the sum of the effect of recognition and the interaction between recognition and tenure is not statistically significant; while the T test for the recognition treatment is positive and statistically significant. A possible explanation for these results is that reps with longer tenure, that is, who have worked for the company for longer periods of time, are less concerned about the possibility of not being hired for future campaigns because they have been hired many times in the past. Reps with longer tenure thus react less to the recognition condition than low tenure reps who look to their field managers for the positive reinforcement provided in the weekly e-mails and for the signal that they are likely to be hired for subsequent campaigns. Reps with longer tenure are more sensitive to the monetary incentives condition. These results are consistent with the implicit incentive argument presented for the recognition condition. Reps who have worked for the company for longer periods of time, who have less need of the signal that they will be hired for subsequent campaigns because they have been hired for so many campaigns in the past, do not react do the recognition condition; their performance is no different from that of the control group. The assumption in this argument (tenure results support the implicit incentive mechanism in the recognition condition) is that the need for acknowledgement does not decrease with tenure. If the assumption does not hold the two interpretations of recognition, implicit incentive or need for acknowledgement, cannot be disentangled with the results from the tenure analysis.

Although the correlation among the demographic variables is low, I nevertheless run the previous models considering all three demographic variables simultaneously. The results are similar to those

reported for the individual analyses presented here: positive interaction between women and monetary incentives; positive interaction between monetary incentives and age as well as between feedback and age; positive interaction between monetary incentives and tenure; and negative interaction between tenure and recognition.

VII. Discussion

The previous analysis shows a positive and statistically significant treatment effect for most of the treatment groups. The estimates in Table 5 of the statistically significant effects are between 8 and 14 pp and are statistically indistinguishable. Concern about a Hawthorne effect is mitigated by the fact that the attention received by Group C (feedback) and, to a lesser extent, Group H (incentives, feedback, and recognition) was similar to that received by the other groups but their performance did not differ from that of the control group. Group H was the group receiving more communications from the company. Concern about a Hawthorne effect is further mitigated by the fact that even reps in the control group received from the field manager two e-mails regarding store goals the company was implementing for the specific manufacturer/retailer/product campaign object of the experiment. Hence, even the control group received some kind of "treatment." The treatments, moreover, were by design implemented by different people in the company in order to separate any person or job position effect from the "real" treatment effect. As described previously, bonus information was sent by a person in the payroll department, the feedback report by someone in the operations department, and the recognition e-mails by the field manager of the rep.

Group C's nonsignificant result can be partially explained by the implementation of the experiment and the research site. Even though the reps received an e-mail describing their treatment at the start of the experiment it was only after four weeks that the weekly e-mails about the reps' bonus, store sales and/or awards were sent. This issue is common to all the groups but may have hurt primarily the feedback group

inasmuch as any learning through feedback could only have occurred in the reporting period. This reduces the duration of the experiment by half, from eight weeks to four. Moreover, the reps can work in one store one week and a different store in the following week which limits the learning process for the feedback group because any given week's report has information only about events the rep performed the prior week. Furthermore, reps were receiving reports for two products that had very different active periods; product A had events mainly in weeks 19, 21, and 22, product B mainly in weeks 19, 24, and 25. But reps only began to receive weekly e-mails on week 23, the report for which had for product B only one week, and for product A two weeks, of store sales. The report sent on week 24 had only information on product A, and the report sent on week 25 information on one week of product B. Most reps were thus receiving feedback on store sales of product A while conducting events for product B, rendering the utility of the feedback questionable. In this study I analyze exclusively data for product B due to the reporting issue and differences in the products' success rates. How feedback interacted with the goals is also an issue. The groups receiving feedback had precise information about the sales value and goals set for each product in the stores in which they worked, whereas groups not receiving feedback, were informed only whether they had earned a bonus and, if so, in what amount, and/or to what award category they had achieved. Hence, the non-feedback groups had no idea how far off they were from the goals in case they did not earn the bonus and/or the award. This situation of missing the goals was quite frequent in the experiment as only 7% of the events of product B met the goal or above, 40% for product A. The conjecture that reps receiving feedback might have been de-motivated by seeing how far they were from the goal is not corroborated by the quantitative evidence in Table 5, which shows performance to have been better during the reporting than during the non-reporting period for the feedback group as well as for most other groups. To shed light on this issue I use evidence from the post-experimental questionnaire. The post-experimental questionnaire was completed by 166 of the 304 reps to whom it was distributed. In untabulated results, I find that respondents and non-respondents do not differ from one another in terms of gender, tenure, or performance. The only statistically significant difference is that respondents are older than non-respondents.

Evidence from the post-experimental questionnaire is mixed in respect to the effect of goal difficulty. Some de-motivation occasioned by the difficult goals set is suggested by qualitative evidence such as the following remarks. “The sales goals for my local store were unrealistic, especially for [product B]. The goal was almost double the actual number of [product B] sold.” “I think that the goals set up at the location are unreachable; please consider that we are at the stores four hours only and usually just make one visit per weekend, so it’s almost impossible to sell [units of product B] or more in four hours. Personally, it demotivates me.” Quantitative evidence, on the other hand, shows that the goals were not rated any more difficult by reps in the feedback groups than by reps in the non-feedback groups. Nor are there statistically significant differences in goal difficulty for the two products in the experiment. On a scale of 1, easiest, to 10, most difficult, the mean for product A was 6.79 and for product B 6.68, and the p-value of the t-test mean comparison 0.45. Relative to the non-feedback groups, the feedback groups did not rate the goals any more difficult, if anything, rated them less difficult, although the differences are not statistically significant. Mean rates for product A were 5.62 for the feedback groups and 7.23 for the non-feedback groups, and for product B, 6.48 for the feedback groups and 6.87 for the non-feedback groups. A two-sided mean comparison found none of the mean differences between the feedback and non-feedback groups to be statistically significant. Hence, the goal difficulty issue is a concern for all groups, not just the feedback groups and hence cannot explain the nonsignificant result for the feedback condition. Overall, I conclude that the nonsignificant result for the feedback treatment is explained neither by negative reporting nor by goal difficulty, leaving the increased perception of control and limited opportunities for learning effects as possible explanations. As one rep stated on the post-experimental questionnaire: “I was aware that the company could see my performance.”

The marginally nonsignificant result for Group H (incentives, feedback, and recognition) in the first specification is puzzling. It might be explained by fewer observations in the experimental period relative to the other groups, although in the initial randomization the number of reps in Group H and the other treatment groups was similar. But fewer observations do not explain the estimated coefficient being lower

than those of other groups. Because reps in this group received a lengthy initial e-mail describing the three treatments, and every week throughout the experiment three different e-mails from three different people, some crowding out effect of the treatments might be occurring. Moreover, the lower estimated coefficient is mainly from the non-reporting period, and the estimated coefficient for the reporting period relative to the pre-experimental period is identical to those of other groups (Table 5, column 4). Reps faced with so many initiatives occurring simultaneously may not have viewed any individual initiative as particularly important, and only when they started receiving the weekly reports realized that the company was actually "serious" about the programs.

Another interesting finding of this field experiment is that the treatment effect is similar for monetary incentives and recognition (the coefficient estimate for recognition is even larger than that for incentives in the first and fourth specifications in Table 5, but the differences are not statistically significant). But the monetary incentives treatment has a direct, albeit not large, cost to the company, whereas the recognition program has none. The bonus program for the two months of the experiment cost $2.410 for 159 reps in the incentive condition. The only cost that can be said to be incurred by the recognition program is that of the time spent by the CEO signing certificates at the end of the experiment (a departure from the original design of the intervention, reps were informed in the kick-off e-mail that the certificates would bear the signatures of their field manager). One possible explanation for the effects of monetary incentives and recognition being identical is that the value of the bonus was too low (per week, $10 extra for reps who met the "Goal," $20 dollars for reps who met the "Stretched Goal," and $30 for reps who met the "Super Stretched Goal"). I address this concern both qualitatively and quantitatively. First, any increase in the bonus scheme implies an increase in the cost of the program, and as there may exist nonlinearities in the relationship between monetary incentives and performance it is an empirical question whether the performance improvement induced by a larger bonus is sufficient to offset the increase in cost. Second, the bonus effect can vary as a function of percentage on pay. The hypothesis is that only monetary incentives that represent a significant proportion of pay have an effect on performance, and hence the

concern that the bonus value is low is eliminated. I interact the treatments with a dummy variable for high number of hours worked each week (using two specifications, more than one event and more than two events in a week), and in untabulated results I find no statistically significant interaction with monetary incentives. The concern about bonus value is further mitigated by a question on the post-experimental questionnaire about the value of the bonus. To the question: "The value of the bonus you could earn was" on a scale from 1, too low, to 10, too high, the mean answer was 5.23 for reps in the incentive condition and 5.76 for reps in Group B (the group that received only the bonus program). But to the question: "What do you think is the reasonable and fair value in dollars of a weekly bonus associated with each level of goals?" the mean answers, for reps in the incentive condition, were $30 dollars for "Goal," $50 for "Stretched Goal," and $77 for "Super Stretched Goal," and, for reps in group B, $20, $40, and $72 dollars, respectively (median values do not differ significantly from mean values). Thus, although they do not consider the values of the bonuses too low, reps' idea of a fair value for the bonuses is in all cases at least twice the experimental amounts.

The effect of the recognition condition can be explained by two different mechanisms, namely the social acknowledgement and the implicit incentive. From an organizational behavior perspective, recognition works as a function of social needs, positive reinforcement, and learning. Because the reps do not even know their peers any effect of public appraisal should be minimal. The close relation between rep and field manager brings into play the reinforcement and acknowledgement conferred by the latter. Within an economic frame, recognition is an implicit incentive that signals future promotion, but this is not the case at the research site because reps are rarely promoted to managers. Recognition can work in this company, however, as a signal that a rep's services will be engaged for future campaigns. I address this issue by including in the post-experimental questionnaire the question: "How would you rate the chance of being hired for the next (manufacturer-retailer) campaign?" The mean answer does not differ between reps in the recognition condition and reps in groups that did not receive this treatment. This result holds even after conditioning on having received a good report throughout the experimental period.

Hence, there is no evidence that the perception of the recognition condition as a signal of future work differed between reps in that condition and reps in groups that received the other treatments. Reps' remarks about the Special Acknowledgement Program were mixed. "Good to have for overall morale." "All I received was an acknowledgement that I hit my goal. That was nice, but I did not receive any compensation. So although I appreciate that they appreciate me, send me a gift card for Starbucks or something!" Despite the evidence from the post-experimental questionnaire does not support the implicit incentive argument, the analysis of the tenure variable is consistent with the implicit incentive argument. Longer tenured reps, being more confident of getting future assignments because they have gotten them in the past, do not react to the recognition treatment (their performance is no different from that of the control group). Shorter tenured reps, on the other hand, react positively to the recognition condition. For this result to be evidence supporting the implicit incentive argument, it is necessary to assume that social need does not vary with tenure. If reps with less tenure in the company look for more reinforcement from, and have a greater need for acknowledgement by, field managers, then the tenure results are consistent with the "social" argument as well. Therefore, I conclude that these alternative mechanisms by which recognition operates cannot be disentangled in this field experiment.

That the effects of monetary incentives and recognition do not add up when considered jointly is an important result since it is evidence that these two treatments are substitutes and not complements. The improved performance of the bundle group (F, incentives and recognition) is less than the sum of the increases in the monetary incentives (B) and recognition (D) group, and this difference is statistically significant. In fact, in Table 5 the estimated effect for Group F is lower than that for Group D (recognition) and/or Group B (incentives), although the difference is not statistically significant. There are two possible explanations for this crowding out effect between the two treatments. There may exist, on one hand, a constraint on how much the treatments influenced reps' effort and motivation that is ultimately reflected in performance. On the other hand, there may exist a constraint on demand that limits sales no matter how much additional effort and motivation reps expend due to the treatments. The latter

explanation seems less plausible than the former given that the estimates of the individual treatment effects are higher than the estimate of the effect for the group that received both treatments, although the difference is not statistically significant. The same pattern of results (the effect of the group that received both monetary incentives and recognition is the same as the effect of the groups that received only one of the treatments and less than the sum of the effects of the groups that received only one treatment) holds when the analysis is done at different deciles of performance, especially for low levels (Table 6). This evidence is important because limitations on demand should not be present at low deciles of performance, which have a larger improvement margin given that the goals were set based on stores' past performance. The estimated treatment effects in the low deciles are in fact larger than the ones in the high deciles, but again the effect of Group F is no different than that of groups B and D and is less than the sum of the effects of groups B and D. Another possible explanation is that ability, training, and experience may limit how much reps can influence sales. This conjecture is not supported in the analysis of heterogeneous treatment effects (gender, age and tenure), in which the negative interaction between monetary incentives and recognition holds. Albeit all this evidence, these alternative explanations cannot be completely ruled out because this field experiment includes neither a separate measure of unconstrained sales nor a measure of effort without limitations. These measures are difficult to obtain in a field setting like retailing in which sales depend on demand as well as reps' effort. In either case, the conclusion in this setting is that there is a substitution effect between monetary incentives and recognition, i.e., the effect of the group that received both treatments does not differ from the effect of the groups that received only one of the treatments, and is less than the sum of the effects of the groups that received only one of the treatments. To assess whether reps preferred one treatment relative to the other I included in the post-experimental questionnaire questions about each intervention. The reps in Group F (incentives and recognition) rated the bonus and special acknowledgement programs similarly in terms of importance, commitment to achieve the bonus/award, and interest in receiving the weekly information. Likewise the effects of the programs in terms of effort, motivation, and learning were also rated similarly by these Reps.

The result that the effects of monetary incentives and recognition do not add up when considered jointly contradicts the interpretation of the recognition condition as an implicit incentive, i.e., if recognition is seen by reps as a signal of a desired future outcome as being re-hired for future campaigns, which translates into hourly wages. If recognition signals money in reps' futures, it is difficult to understand why monetary incentives would work only up to the point of one week's bonus of $10-$30 (explicit), but not for the added hourly wages that would come from future assignments (implicit). A utility function that flattens after the $30 bonus seems implausible.

For the other interactions that involve the feedback condition and are not statistically significant, I cannot claim independence of the treatments because the feedback condition is not statistically significant by itself and, by definition, a non-effect (feedback) is independent of any statistically significant effect (monetary incentives and recognition).

VIII. Conclusion

This study provides experimental evidence from a corporate setting regarding the performance effects of providing monetary incentives, performance feedback, and recognition to lower level employees. Using as a research site a retail services company, I find that sales reps who received only monetary incentives increased their performance by 8 percentage points, which contrasts with some laboratory evidence (Bonner et al. 2000) but is consistent with prior evidence from archival (Prendergast 1999) and field studies (Banker et al. 1996). Sales reps who received only recognition increased their performance by 12 percentage points, which is consistent with psychology theory on the need for acknowledgement (Maslow 1943, McClelland 1961, Alderfer 1972) and the signaling mechanism (Bandura 1999), which is also discussed in economics (Bénabou and Tirole 2003). The introduction of performance feedback does not have a statistically significant effect on performance, which contrasts with the evidence from goal theory (Locke and Latham 1990). To address the call of Holmstrom and Milgrom (1994) that different rewards be analyzed as part of a coherent incentive system, I compare the effects of monetary incentives, feedback, and recognition and find the effects of monetary incentives and recognition to not be

statistically different from one another (both are different from feedback, but feedback is not statistically significant by itself). Moreover, using the definition of complements from Milgrom and Roberts (1995), I find the increase in performance for the group that received both monetary incentives and recognition to be 11 percentage points below the sum of the performance increases for the groups that received only one of the treatments. Hence, I find evidence that monetary incentives and recognition are substitutes in this setting. I argue that limitations on the demand side and in the ability and training of sales reps that might account for why monetary incentives and recognition do not add up do not explain my results because they are robust to the reps' performance level, working experience, and tenure.

These treatment effects are heterogeneous in terms of the deciles of the performance distribution and demographic characteristics of the sales reps. I find that the treatments are most effective in the bottom deciles of the performance distribution, and that demographic variables, namely gender, age, and tenure, influence the effects of the different treatment conditions. I find that monetary incentives are effective for women but not for men, and vice versa for recognition. Older (more experienced) reps react more to monetary incentives and performance feedback and less to recognition, and vice versa for younger (less experienced) reps. High tenure reps react positively to monetary incentives, which is not the case for low tenure reps, whereas recognition has a positive effect on the performance of low tenure reps but does not affect the performance of high tenure reps.

This study makes a number of contributions. First, I show that it is important to consider extrinsic rewards other than monetary incentives, which have been the exclusive component of the majority of the principal agent models used in accounting research. I find evidence that recognition can be as effective as monetary incentives without incurring the added direct monetary cost of a bonus. Second, I show that it is important to consider in an integrated perspective monetary incentives, feedback, and recognition in decisions related to designing a reward system. I find evidence that monetary incentives and recognition are substitutes. Although this result may be constrained somewhat by limitations on demand and in the ability and training of reps, these boundary conditions are common to most settings in the corporate

world. Third, I show that, there being important heterogeneous treatment effects, the design of incentive systems should take into account the diversity of the workforce. I find evidence that the performance effects of monetary incentives, feedback, and recognition vary with reps' performance level, gender, age, and tenure. Hence, the "one size fits all" rule seems not to apply to the design of incentive systems. Overall, I contribute to the literature by providing experimental evidence from the corporate world on the importance of having an integrated perspective on extrinsic rewards and taking into consideration the characteristics of a diverse work force when designing incentive systems.

This study has a number of implications for practice. First, I show that recognition can be as effective as monetary rewards that have an added direct cost. Second, I show that feedback may have no effect on performance and thus companies should evaluate the usefulness of reports provided to their employees. Third, I show that introducing multiple extrinsic rewards is no different from introducing just one (monetary incentives or recognition) in terms of performance effects. Hence, companies should not waste resources on multiple extrinsic reward programs. Fourth, I show that extrinsic reward programs are more effective at lower levels of performance, and that other characteristics of the employees, namely demographics, also affect the success of reward programs.

This study is not without caveats. The limitations are related mainly to the method used and the research site. Any field study, and thus any field experiment, faces concerns about the generalizability of results. I argue that the conclusions drawn here should be applicable to other settings in which the control of front-line employees is loose and interaction among them is low. If anything, this low interaction environment should yield a conservative estimate of the treatment effects inasmuch as it lacks the peer pressure and peer comparison to which employees in most organizations are subjected. Nevertheless, this setting enabled me to run a field experiment with multiple treatment groups and virtually no contamination, thus increasing the internal validity of the study. It further afforded me the opportunity to test in the corporate world, but with the rigor of an experiment, the performance effects of introducing different extrinsic rewards, which is an advantage relative to both archival studies and laboratory

experiments. The substitution effect identified can also be a product of the setting owing to limits on increasing sales occasioned by demand constraints and the ability and experience of sales reps. I mitigate these concerns with additional analyses, and further argue that, given that the corporate world lives with limitations on demand and available resources, the fact that the effects of monetary incentives and recognition do not add up should be generalizable to other settings. A similar argument is valid for heterogeneous treatment effects. The diversity of its workforce makes the collaborating research site an ideal one in which to test the heterogeneity of the treatment effects. Findings related to the heterogeneous treatment effects are applicable even to companies with more homogeneous workforces, which can choose the rewards that best suit their employees. Another limitation is the short period of data available; the treatment effects were estimated with only two months of data and one month of reporting. Nevertheless, compared to a laboratory experiment with a one or multi-period manipulation that lasts, at most, a couple of hours, my field experiment has a long window and a "real" setting. I also addressed this concern empirically by using cutoffs in the data, which revealed the performance effects to usually be stronger in the reporting than in the non-reporting period.

These caveats and the puzzling questions presented in the discussion of the results are important, and represent multiple avenues for future research. For example, disentangling the implicit incentive from the social acknowledgement component in the recognition condition would contribute to the design of better incentive systems. Whether the results hold in settings involving peer pressure and frequent interaction among employees would be another interesting line of inquiry. Finally, whether the treatment effects will persist over a longer experimental window is an important question for both academics and practitioners given the long horizon of most contractual relationships in the workplace. It would be interesting to test as well whether the treatment effects persist after the rewards are removed. Answers to these questions would help academics fine-tune their theories on contracts and contribute to the design of more successful and efficient reward programs that would benefit practitioners. This study is necessarily a starting point in such promising new lines of research.

References

Alderfer, C. P. (1972). *Existence, Relatedness, and Growth; Human Needs in Organizational Settings*. New York: Free Press.

Angrist, J. D. and A. B. Krueger. 2001. Instrumental Variables and Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives* 15 (4): 69-85.

Arceneaux, K., I. Gerber and D. Green. 2006. Comparing Experimental and Matching Methods using a Large-Scale Field Experiment on Voter Mobilization. *Political Analysis* 14 (1): 37-62.

Ariely, D., A. Bracha, and S. Meier. 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review* 99 (1): 544-555.

Baiman, S. (1990). Agency research in managerial accounting: a second look. *Accounting, Organizations and Society* 15 (4): 341–371.

Baker, G. P., M. C. Jensen and K. J. Murphy. 1988. Compensation and incentives: Practice vs. Theory. *Journal of Finance* 43 (3): 593-616.

Baker, G. P., R. Gibbons and K. J. Murphy. 1994. Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics* 109 (4): 1125-1156.

Bandiera, O., I. Barankay, and I. Rasul. 2005. Social preferences and the response to incentives: Evidence from personnel data. *Quarterly Journal of Economics* 120 (3): 917-962.

Bandura, A. 1986. *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. 1991. Social cognitive theory of self regulation. *Organizational Behavior and Human Decision Processes* 50(2): 248–287.

Bandura, A. 1999. Social cognitive theory of personality in L. Pervin & O. John (Eds), *Handbook of personality* (2nd ed.): 154-196. New York: Guilford.

Banker, R. D. and S. M. Datar. 1989. Sensitivity, precision, and linear aggregation of signals of performance evaluation. *Journal of Accounting Research* 27 (1): 21-39.

Banker, R. D., L. Seok-Young and G. Potter. 1996. A field study of the impact of a performance-based incentive plan. *Journal of Accounting and Economics* 21 (2): 195 226.

Bartling, B., E. Fehr, and K. M. Schmidt. 2010. Screening, competition, and job design. University of Zürich, Institute for Empirical Research in Economics, Working Paper No. 470.

Bénabou, R. and J. Tirole. 2003. Intrinsic and extrinsic Motivation. *The Review of Economic Studies* 70 (3): 489-520.

Bénabou, R. and J. Tirole. 2006. Incentives and Prosocial Behavior. *The American Economic Review* 96 (5): 1652-1678.

Bonner, S. E., R. Hastie, G. B. Sprinkle, and S. M. Young. 2000. A review of the effects of financial incentives on performance by laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research* 12(1): 19–64.

Cameron, J. and W. D. Pierce. 1994. Reinforcement, reward, and intrinsic motivation: a meta-analysis. *Review of Educational Research* 64 (3): 363-423.

Cameron, J., K. M. Banko, and W. D. Pierce. 2001. Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst* 24 (1): 1-44.

Campbell, D. 2008. Nonfinancial Performance Measures and Promotion-Based Incentives. *Journal of Accounting Research* 46 (2): 297-332.

Chen, Y. 2000. Promises, Trust and Contracts. *Journal of Law, Economics, and Organizations* 16 (1): 209-232.

Deci, E. L. 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18 (1): 105-115.

Deci, E. L. 1972. Intrinsic motivation, extrinsic reinforcement and inequity. *Journal of Personality and Social Psychology* 22 (1): 113-120.

Deci, E. L., R. Koestner and R. M. Ryan. 1999. A Meta-analytical review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125 (6): 627-668.

Deci, E., L. 1975. *Intrinsic motivation*. New York; London: Plenum Press.

Deci, E.L. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.

Eisenberg, R. and J. Cameron. 1996. Detrimental effects of reward: Reality or myth? *American Psychologist* 51 (11): 1153-1166.

Fehr, E. and J. A. List. 2004. The Hidden costs and returns of incentives – Trust and trustworthiness among CEOs. *Journal of the European Economic Association* 2 (5): 743-771.

Fehr, E. and S. Gatcher. 2000. Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives* 14 (3): 159-181.

Fehr, E., A. Klein, and K. M. Schmidt. 2007. Fairness and contract design. *Econometrica* 75 (1): 121-154.

Feltham, G. A. and J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69 (3): 429-453.

Frey, B. S. 1994. How intrinsic motivation is crowded out and in. *Rationality and Society* 6 (3): 334-352.

Frey, B. S. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham, UK/Brookfield, USA: Edward Elgar.

Frey, B. S. 2006. Giving and receiving awards. *Perspectives on Psychological Science* 1 (4): 377-388.

Frey, B. S. 2007. Awards as compensation. *European Management Review* 4 (1): 6-14.

Frey, B.S. and R. Jegen. 2001. Motivation crowding theory: A survey of empirical evidence. *Journal of Economic Surveys* 15 (5): 589-611.

Gibbons, R. 1998. Incentives in Organizations. *Journal of Economic Perspectives* 12: 115-132.

Glaeser, E. L., D. I. Laibson, J. A. Scheinkman, and C. L. Soutter. 2000. Measuring trust. *The Quarterly Journal of Economics* 115 (3): 811-846.

Hannan, L., R. Krishnan, and A. Newman. 2008. The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *The Accounting Review* 83 (4): 893-913.

Harrison, G. and J. List. 2004. Field Experiments. *Journal of Economic Literature* XLII: 1013-1059.

Holmstrom, B. 1979. Moral Hazard and Observability. *Bell Journal of Economics* 10: 74-91.

Holmstrom, B. and P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organization* 7:24-52.

Holmstrom, B. and P. Milgrom. 1994. The firm as an incentive system. *The American Economic Review* 84 (4):972-991.

Ichniowski, C., K. Saw, and G. Prennushi. 1997. The effects of human resource management practices on productivity: a study of steel finishing lines. *The American Economic Review* 87 (3): 291-313.

Ittner, C. D. and D. F. Larcker. 2001. Assessing empirical research in managerial accounting: a value-based management perspective. *Journal of Accounting and Economics* 32 (1-3): 349-410.

Jensen, M. C. and W. H. Meckling. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. Journal of Financial Economics 3 (4): 305-360.

Kachelmeier, S. J. and R. R. King. 2002. Using Laboratory Experiments to Evaluate Accounting Policy Issues. *Accounting Horizons* 16 (3): 219-232.

Kerr, S. 1975. On the folly of rewarding for A, while hoping for B. *The Academy of Management Journal* 18 (4): 769-783.

Kluger, A. N., and A. DeNisi. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119 (2): 254–284.

LaLonde, R. J. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *American Economic Review* 76 (4): 604-620.

Latham, G. P. and E. A. Locke. 1991. Self-Regulation through Goal Setting. *Organizational Behavior and Human Decision Processes* 50 (2): 212-247

Lazear, E. 1986. Salaries and Piece Rates. *Journal of Business* 59 (3): 405-431.

Lazear, E.P. 1999. Personnel economics. Past lessons and future directions. *Journal of Labor Economics* 17(2): 199-236.

Lepper, M. R., D. Greene and R. E. Nisbet. 1973. Undermining children's intrinsic interest with extrinsic reward: a test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology* 28 (1): 129-137.

Lepper, M.R. and D. Greene. 1978. *The hidden costs of reward: new perspectives on the psychology of human motivation*. New York: Erlbaum.

Levitt, S. and J. List. 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives* 21 (2): 153-174.

Levitt, S. and J. List. 2009. Field Experiments in Economics: the Past, the Present and the Future. *European Economic Review* 53(1): 1-18.

Locke, E. A., and G. P. Latham. 1990. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.

Luckett, P. F. and I. R. Eggleton. 1991. Feedback and Management Accounting: a Review of Research into Behavioural Consequences. *Accounting Organizations and Society* 16 (4): 371-394.

MacLeod, W. B. and D. Parent. 1999. Job characteristics and the form of compensation. *Research in Labor Economics* 18: 177-242.

Maslow, A. H. 1943. A Theory of Human Motivation. *Psychological Review* 50 (4): 370-396.

Mcclelland, D. C. 1961. *The achieving society.* Princeton, N. J.: Van Nostrand.

Milgrom, P. and J. Roberts. 1990. The economics of modern manufacturing: technology, strategy and organization. *The American Economic Review* 80 (3): 511-528.

Milgrom, P. and J. Roberts. 1995. Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics* 19 (2-3): 179-208.

Neckerman, S. and Frey, B. S. 2008. Awards as compensation. CREMA working paper 2008-31.

Ostrom, E. 2000. Collective action and the evolution of social norms. *The Journal of Economic Perspectives* 14 (3): 137-158.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1): 7–63.

Rabin, Mathew. 1993. Incorporating fairness into game theory. *American Economic Review* 83: 1281-1302.

Skinner, B. F. 1953. *Science and human behavior*. New York: Macmillan.

Sprinkle, G. B. 2000. The Effect of Incentive Contracts on Learning and Performance. *The Accounting Review* 75 (3): 299-326.

Sprinkle, G. B. 2003. Perspectives on Experimental Research in Managerial Accounting. *Accounting, Organizations and Society* 28 (2-3): 287-318.

Vroom, V. H. 1964. *Work Motivation*. New York: Wiley.

Figure 1 – Experimental design

Treatments		Incentives			
		No		Yes	
		Recognition		Recognition	
		No	Yes	No	Yes
Feedback	No	Control Group A	Recognition Group D	Incentives Group B	Incentives & Recognition Group F
	Yes	Feedback Group C	Feedback & Recognition Group G	Incentives & Feedback Group E	Incentives & Feedback & Recognition Group H

Figure 2 – Experimental groups

Groups \ Treatments	Incentives	Feedback	Recognition
Group A – Control	Control (no change)		
Group B – Incentives	x		
Group C – Feedback		x	
Group D – Recognition			x
Group E – Incentives and Feedback	x	x	
Group F – Incentives and Recognition	x		x
Group G – Feedback and Recognition		x	x
Group H – Incentives, Feedback and Recognition	x	x	x

Figure 3 – Timeline of the field experiment

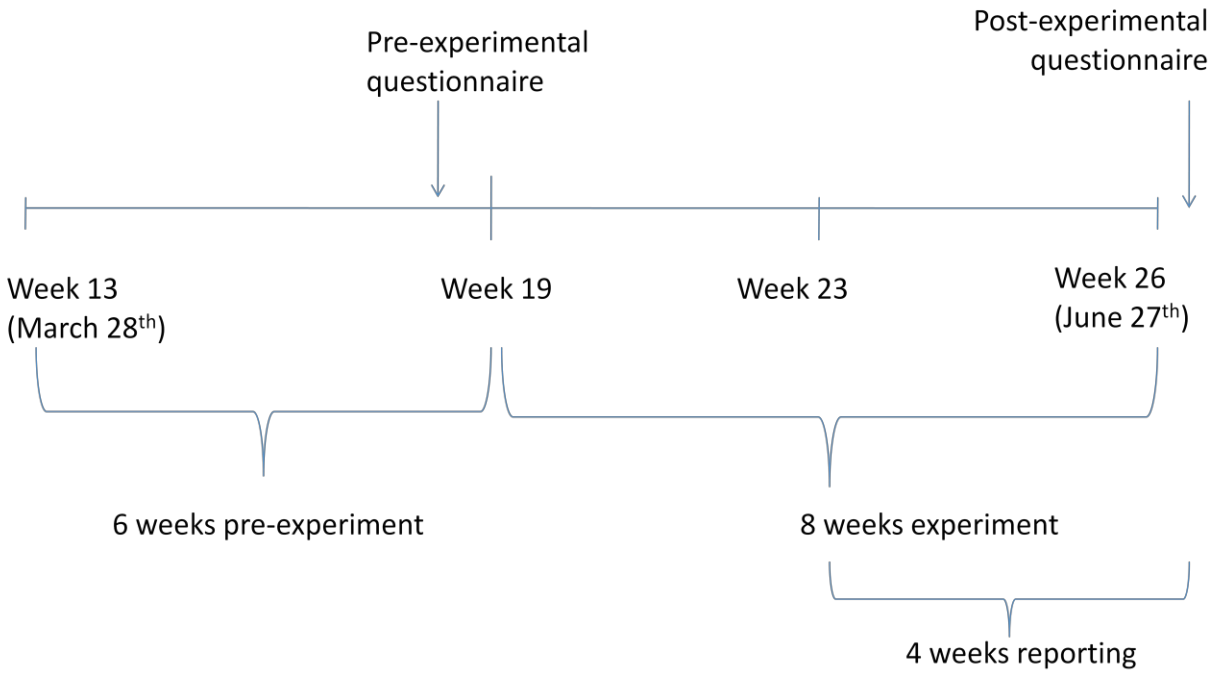

Figure 4 – Number of reps working each week per group

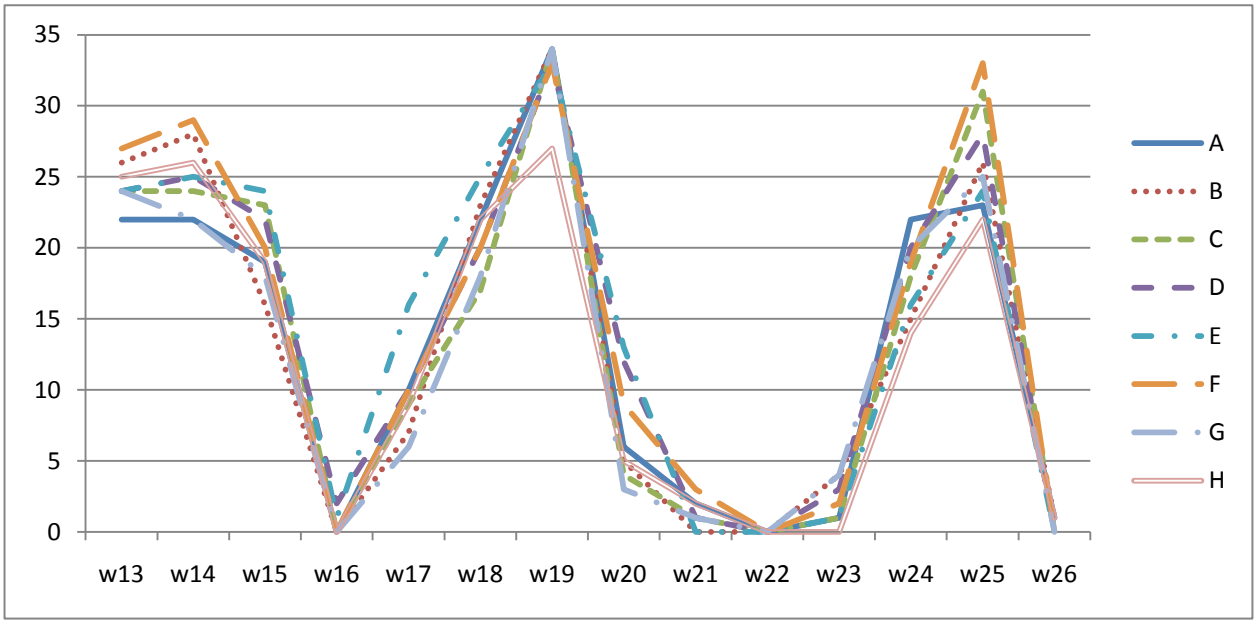

Figure 5 – Mean performance of the groups per week

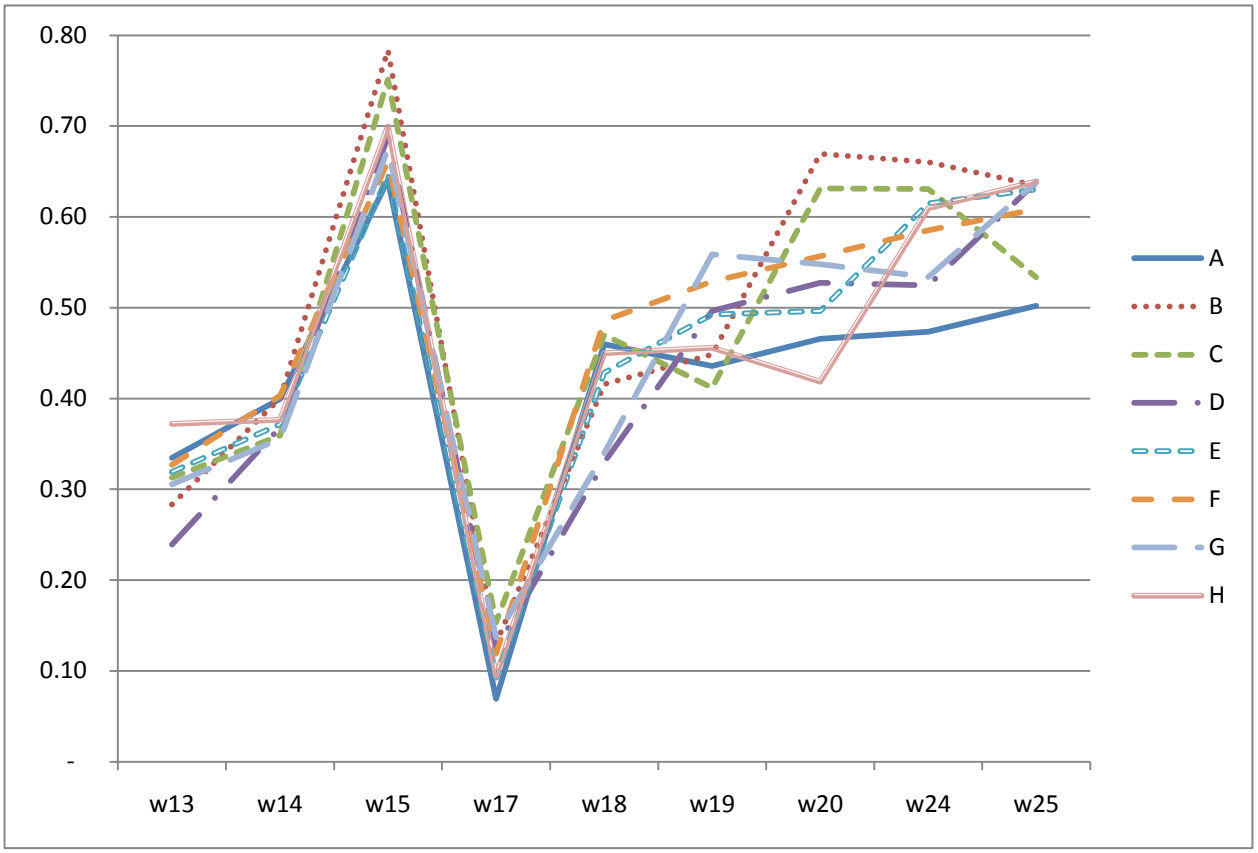

Table 1 – Number of reps per group

Data / Groups	Panel A		Panel B			
	Reps initial assignment	Reps with data	Reps with data for product B	Reps with pre-experimental data for product B	Reps with experimental data for product B	Reps with pre-experimental and experimental data for product B
A	55	45	41	38	36	33
B	55	51	47	40	39	32
C	55	47	45	39	41	35
D	55	50	44	40	38	34
E	55	46	45	43	38	36
F	54	50	49	41	40	32
G	54	52	46	35	40	29
H	54	46	42	40	32	30
X		85	59	2	57	
Total	437	472	418	318	361	261

Table 2 – Characteristics of the groups

Data / Groups	% Female	Tenure in months Mean (Standard Error)	Age in years Mean (Standard Error)
A	.36	8.9 (1.4)	43.0 (2.4)
B	.55	13.2 (2.4)	45.7 (2.0)
C	.4	13.5 (3.0)	41.2 (2.0)
D	.48	12.6 (2.2)	42.2 (1.7)
E	.48	10.8 (1.6)	43.5 (1.8)
F	.48	8.5 (1.4)	42.5 (2.3)
G	.5	10.7 (1.8)	45.2 (2.1)
H	.57	12.0 (2.5)	42.8 (1.9)
X	.55	17.7 (2.5)	46.4 (1.5)
Chi-square p-value a)	0.5	0.45[1)]	0.637[2)]
F stat p-value a)		0.58	0.78

a) Excluding group X, which was not randomized.
1) The test used the dichotomous variable above median tenure.
2) The test used the dichotomous variable above median age.

Table 3 – Descriptive statistics for the outcome variable

Period / (Weeks) Descriptive statistics	All (weeks 13-26)	Pre-experimental period (weeks 13-18)	Experimental period (weeks 19-26)	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Maximum	1.61	1.14	1.61	0.71	0.90	1.14	0.50	0.38	0.93	1.33	1.11	0.70		0.88	1.31	1.62	0.40
Third quartile	0.61	0.54	0.66	0.41	0.48	0.80	0.49	0.15	0.54	0.60	0.65	0.50		0.37	0.71	0.76	0.30
Median	0.44	0.38	0.50	0.32	0.38	0.71	0.48	0.09	0.41	0.45	0.50	0.42		0.33	0.57	0.57	0.20
Mean	0.46	0.40	0.53	0.31	0.38	0.69	0.44	0.11	0.42	0.48	0.53	0.42		0.35	0.57	0.60	0.25
First quartile	0.29	0.24	0.36	0.22	0.27	0.57	0.40	0.06	0.30	0.33	0.40	0.30		0.25	0.42	0.44	0.18
Standard deviation	0.24	0.23	0.23	0.14	0.15	0.19	0.09	0.08	0.20	0.21	0.20	0.17		0.17	0.23	0.26	0.13
Minimum	0.00	0.00	0.00	0.00	0.04	0.23	0.33	0.00	0.00	0.00	0.15	0.15		0.13	0.00	0.07	0.15
Number of observations	1509	805	704	196	201	161	3	77	167	262	57	10	0	16	144	212	3

Table 4 – Means per group of the outcome variable

Period (Weeks) / Groups	All (weeks 13-26)	Pre-experimental period (weeks 13-18)	Experimental period (weeks 19-26)	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	.43	.41	.46	0.33	0.40	0.64		0.07	0.46	0.44	0.47	0.37		0.36	0.47	0.50	0.20
B	.47	.41	.54	0.28	0.40	0.78		0.13	0.42	0.45	0.67			0.23	0.66	0.63	0.40
C	.47	.44	.50	0.31	0.36	0.75		0.15	0.47	0.41	0.63	0.44		0.33	0.63	0.53	
D	.45	.38	.54	0.24	0.37	0.69	0.49	0.12	0.33	0.50	0.53	0.15		0.32	0.52	0.64	
E	.46	.39	.55	0.32	0.37	0.65	0.33	0.09	0.43	0.49	0.50			0.36	0.61	0.63	
F	.49	.42	.56	0.33	0.40	0.66		0.12	0.49	0.53	0.56	0.46		0.56	0.59	0.61	
G	.47	.38	.56	0.31	0.36	0.67		0.14	0.34	0.56	0.55	0.35		0.39	0.53	0.64	
H	.47	.42	.54	0.37	0.38	0.70		0.09	0.45	0.46	0.42	0.55			0.61	0.64	0.15

Table 5 – Treatment effects: Pooled OLS regression for the pre-experimental and experimental periods with clustered standard errors by rep [1)]

Dependent variable: *Average sales/goals per rep-week*	(1) Sample: All periods		(2) Sample: All except week 19		(3) Sample: All except reporting period		(4) Sample: All except non-reporting period	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
B (incentives)	.009	.022	.009	.022	.008	.022	.008	.022
C (feedback)	.018	.024	.018	.024	.017	.024	.017	.024
D (recognition)	-.041	.026	-.041	.026	-.041	.026	-.041	.026
E (incentives & feedback)	-.016	.022	-.016	.022	-.016	.022	-.016	.022
F (incentives & recognition)	.013	.025	.013	.025	.013	.025	.013	.025
G (feedback & recognition)	-.030	.027	-.030	.027	-.030	.027	-.030	.027
H (incentives & feedback & recognition)	.016	.021	.016	.021	.016	.021	.016	.021
B * study period	.082**	.039	.133***	.043	.028	.056		
C * study period	.022	039	.061	.045	-.019	.052		
D * study period	.117***	.040	.127***	.042	.093*	.054		
E * study period	.105**	.044	.125**	.051	.062	.055		
F * study period	.089**	.041	.097**	.046	.078	.055		
G * study period	.138***	.044	.127**	.051	.147***	.061		
H * study period	.061	.038	.096**	.043	.003	.052		
B * study period * reporting period							.128***	.044
C * study period * reporting period							.057	.050
D * study period * reporting period							.137***	.048
E * study period * reporting period							.146***	.059
F * study period * reporting period							.100*	.051
G * study period * reporting period							.133***	.056
H * study period * reporting period							.114**	.050
FE by week	Yes		Yes		Yes		Yes	
F stats treatment groups	1.59		1.59		1.59		1.59	
F stats treatment groups * study period	2.52**		2.27**		2.43**			
F stats treatment groups * study * reporting							2.08**	
F stats week dummies	102.12***		111.2***		146.62***		133.65***	
F stats B+C=E [2)]	0.00		0.94		0.53		0.26	
F stats B+D=F [2)]	3.70*		6.19**		0.35		5.46**	
F stats C+D=G [2)]	0.00		0.78		0.95		0.63	
F stats B+C+D=H [2)]	4.13**		6.61**		0.86		4.64**	
N	1509		1247		1134		1180	
Reps	359		352		340		351	
Adjusted R-squared	.34		.39		.39		.40	

[1)] Two-sided tests, * significant at 10% level, ** significant at 5% level, *** significant at 1% level. Variables B to H are dummy variables for the treatment groups according to the random assignment of the sales reps. Study period is a dummy variable for the eight weeks of the experimental period, and reporting period is a dummy variable for the four weeks of reporting within the experimental period.

[2)] All coefficients refer to the study period except in column 4, in which the coefficients refer to the reporting period.

Table 6 – Quantile treatment effects: Quantile regressions by deciles of performance with bootstrap standard errors by rep with 100 repetitions[1)]

	Dependent variable: Average sales/goals per rep-week								
Decile	10%	20%	30%	40%	50%	60%	70%	80%	90%
B * study period	.134**	.085	.041	.045	.057	.076	.089	.058	.006
	(.059)	(.053)	(.047)	(.050)	(.050)	(.063)	(.062)	(.073)	(.070)
C * study period	.035	.005	.043	.045	.019	.027	.018	-.020	-.025
	(.073)	(.055)	(.045)	(.050)	(.046)	(.051)	(.056)	(.067)	(.090)
D * study period	.142**	.128**	.14***	.1**	.094**	.076	.072	.072	.037
	(.065)	(.062)	(.045)	(.041)	(.041)	(.061)	(.063)	(.060)	(.052)
E * study period	.133**	.058	.1**	.095*	.128**	.122**	.056	.047	.030
	(.056)	(.054)	(.046)	(.053)	(.054)	(.053)	(.048)	(.076)	(.081)
F * study period	.174***	.078	.08**	.054	.056	.065	.076	.027	.003
	(.058)	(.049)	(.041)	(.056)	(.052)	(.066)	(.057)	(.070)	(.106)
G * study period	.199***	.1**	.1**	.077	.091	.112	.108	.102	.083
	(.070)	(.049)	(.048)	(.054)	(.062)	(.078)	(.071)	(.075)	(.107)
H * study period	.072	.037	.079	.056	.085*	.083	.07	.017	-.059
	(.063)	(.052)	(.053)	(.055)	(.051)	(.061)	(.065)	(.071)	(.059)
FE by week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE by group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1509	1509	1509	1509	1509	1509	1509	1509	1509
Reps	359	359	359	359	359	359	359	359	359

[1)] The table shows the estimates of the coefficients and standard errors (in parentheses). Significance levels are shown with * for the 10% level, ** for the 5% level, *** for the 1% level. All tests are two sided. Variables B to H are dummy variables for the treatment groups according to the random assignment of the sales reps. Study period is a dummy variable for the eight weeks of the experiment.

Table 7 – Descriptive statistics for the demographic variables

Panel A – Reps N = 359				
Demographic variables	Mean	Std deviation	Minimum	Maximum
Female	.47			
Age (years)	43.18	14.00	18	82
Tenure (months)	11.18	14.44	2	108
Panel B – Rep Weeks N = 1509				
Demographic variables	Mean	Std deviation	Minimum	Maximum
Female	.42			
Age (years)	43.16	14.05	18	82
Tenure (months)	10.43	13.09	2	108

Table 8 – Extrinsic rewards and gender: Pooled OLS regression for the pre-experimental and experimental periods with clustered SE by rep[1)]

Dependent variable: *Average sales/goals per rep-week*	(2) All treatments together		(3) All treatment groups separated		(4) Just main treatments		(5) Effect coding	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Female	.001	.012	-.037	.032	-.021	.024	-.003	.012
Female * study period	-.106*	.063	-.069	.066	-.011	.044	-.018	.021
Treatment	.054*	.030						
Treatment * female	.101	.066						
B (Incentives) * study period			.043	.044	-.015	.027	-.010	.014
C (Feedback) * study period			-.0004	.050	.033	.028	.016	.014
D (Recognition) * study period			.124***	.044	.067**	.027	.033**	.014
E (Incentives & Feedback) * study period			.081	.056			.004	.014
F (Incentives & Recognition) * study period			.022	.045			-.042***	.014
G (Feedback & Recognition) * study period			.173***	.056			.007	.014
H (Incentives & Feedback & Recognition) * study period			.067	.049			-.005	.014
B (Incentives) * study period * female			.099	.085	.081*	.043	.046**	.021
C (Feedback) * study period * female			.074	.084	-.049	.043	-.019	.021
D (Recognition) * study period * female			-.001	.090	-.036	.043	-.016	.021
E (Incentives & Feedback) * study period * female			.096	.094			-.024	.021
F (Incentives & Recognition) * study period * female			.182**	.089			.016	.021
G (Feedback & Recognition) * study period * female			-.053	.093			-.037*	.021
H (Incentives & Feedback & Recognition) * stdyprd * female			.013	.084			-.005	.021
FE by week	Yes		Yes		Yes		Yes	
FE by group/treatment	No		Yes		Yes		Yes	
FE by group/treatment * female	No		Yes		Yes		Yes	
F stats total effect (treatment + treatment * female)	6.99*							
Incentives					3.82*		4.90**	
Feedback					0.22		0.02	
Recognition					0.90		1.05	
N	1509		1509		1509		1509	
Reps	359		359		359		359	
Adjusted R-squared	.35		.35		.34		.35	

[1)] Two-sided tests, * significant at 10% level, ** significant at 5% level, *** significant at 1% level. In columns 2 and 3, the treatment groups, variables B to H, were coded as dummy variables according to the random assignment of the sales reps. In column 4, the treatments, monetary incentives, feedback, and recognition, were coded as dummy variables according to the random assignment of the sales reps and interactions among the treatments ignored to avoid multicollinearity. In column 5, effect coding implied the value of +1 if the rep was provided with a given treatment condition and the value of -1 if the rep was not subject to that condition, according to the random assignment of the sales reps and subsequent calculation of the interactions among the treatments. In all columns, study period is a dummy variable for the eight weeks of the experimental period, and female is the dummy variable for gender.

Table 9 – Extrinsic rewards and age: Pooled OLS regression for the pre-experimental and experimental periods with clustered SE by rep[1)]

Dependent variable: *Average sales/goals per rep-week*	(2) All treatments together		(3) All treatment groups separated		(4) Just main treatments		(5) Effect coding	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Age	-.007	.006	-.031**	.012	-.019*	.010	-.007	.006
Age * study period	-.037*	.022	-.014	.025	-.017	.018	.021**	.010
Treatment	.085***	.028						
Treatment * age	.072***	.024						
B (Incentives) * study period			.078**	.040	.019	.021	.009	.010
C (Feedback) * study period			.023	.040	.013	.021	.002	.010
D (Recognition) * study period			.114***	.040	.049**	.021	.022**	.010
E (Incentives & Feedback) * study period			.101**	.043			-.004	.010
F (Incentives & Recognition) * study period			.087**	.040			-.030***	.010
G (Feedback & Recognition) * study period			.116***	.042			-.009	.010
H (Incentives & Feedback & Recognition) * study period			.059	.039			-.004	.010
B (Incentives) * study period * age			.017	.039	.041**	.020	.016	.010
C (Feedback) * study period * age			.050	.036	.053***	.020	.026**	.010
D (Recognition) * study period * age			-.042	.038	-.001	.020	-.006	.010
E (Incentives & Feedback) * study period * age			.099**	.044			-.015	.010
F (Incentives & Recognition) * study period * age			.064**	.031			-.001	.010
G (Feedback & Recognition) * study period * age			.069**	.039			-.007	.010
H (Incentives & Feedback & Recognition) * stdyprd * age			.027	.038			-.022**	.010
FE by week	Yes		Yes		Yes		Yes	
FE by group/treatment	No		Yes		Yes		Yes	
FE by group/treatment * age	No		Yes		Yes		Yes	
F stats total effect (treatment + treatment * age)	14.80***							
Incentives					4.17**		3.03*	
Feedback					5.03**		3.72*	
Recognition					2.72*		1.18	
N	1509		1509		1509		1509	
Reps	359		359		359		359	
Adjusted R-squared	.35		.36		.35		.36	

[1)] Two-sided tests, * significant at 10% level, ** significant at 5% level, *** significant at 1% level. In columns 2 and 3, the treatment groups, variables B to H, were coded as dummy variables according to the random assignment of the sales reps. In column 4, the treatments, monetary incentives, feedback, and recognition, were coded as dummy variables according to the random assignment of the sales reps and interactions among the treatments ignored to avoid multicollinearity. In column 5, effect coding implied the value of +1 if the rep was provided with a given treatment condition and the value of -1 if the rep was not subject to that condition, according to the random assignment of the sales reps and subsequent calculation of the interactions among the treatments. In all columns, study period is a dummy variable for the eight weeks of the experimental period, and age is a standardized measure of the number of years-old of each sales rep.

Table 10 – Extrinsic rewards and tenure: Pooled OLS regression for the pre-experimental and experimental periods with clustered SE by rep[1)]

Dependent variable: *Average sales/goals per rep-week*	(2) All treatments together		(3) All treatment groups separated		(4) Just main treatments		(5) Effect coding	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Tenure	-.007	.005	-.043**	.022	-.015**	.007	-.014***	.004
Tenure * study period	.013	.032	.049	.031	-.002	.021	.010	.010
Treatment	.082***	.027						
Treatment * tenure	-.012	.034						
B (Incentives) * study period			.074*	.039	.017	.021	.008	.010
C (Feedback) * study period			.013	.039	.012	.021	.001	.010
D (Recognition) * study period			.111***	.040	.051**	.021	.021**	.010
E (Incentives & Feedback) * study period			.102**	.042			-.004	.010
F (Incentives & Recognition) * study period			.087**	.039			-.033***	.010
G (Feedback & Recognition) * study period			.116***	.041			-.010	.010
H (Incentives & Feedback & Recognition) * study period			.046	.039			-.008	.010
B (Incentives) * study period * tenure			-.046	.034	.037*	.021	.022**	.010
C (Feedback) * study period * tenure			-.043	.033	.002	.022	-.007	.010
D (Recognition) * study period * tenure			-.079**	.036	-.024	.021	-.035***	.010
E (Incentives & Feedback) * study period * tenure			.070	.049			.015	.010
F (Incentives & Recognition) * study period * tenure			-.002	.035			.005	.010
G (Feedback & Recognition) * study period * tenure			-.123***	.044			-.026**	.010
H (Incentives & Feedback & Recognition) * stdyprd * tenure			-.091	.057			-.025**	.010
FE by week	Yes		Yes		Yes		Yes	
FE by group/treatment	No		Yes		Yes		Yes	
FE by group/treatment * tenure	No		Yes		Yes		Yes	
F stats total effect (treatment + treatment * tenure)	3.25*							
Incentives					3.28*		4.78**	
Feedback					0.25		0.23	
Recognition					0.84		0.92	
N	1509		1509		1509		1509	
Reps	359		359		359		359	
Adjusted R-squared	.35		.35		.35		.35	

[1)] Two-=sided tests, * significant at 10% level, ** significant at 5% level, *** significant at 1% level. In columns 2 and 3, the treatment groups, variables B to H, were coded as dummy variables according to the random assignment of the sales reps. In column 4, the treatments, monetary incentives, feedback, and recognition, were coded as dummy variables according to the random assignment of the sales reps and interactions among the treatments ignored to avoid multicollinearity. In column 5, effect coding implied the value of +1 if the rep was provided with a given treatment condition and the value of -1 if the rep was not subject to that condition, according to the random assignment of the sales reps and subsequent calculation of the interactions among the treatments. In all columns, study period is a dummy variable for the eight weeks of the experimental period, and tenure is a standardized measure of the number of months the sales rep has been working for the retail services company.

Appendix 1 – Initial communication

Control

"Hello,

For the XXX events in the month of May, the management has decided to set weekly sales goals for every store.
Please locate the goal figures of the stores for which you are scheduled in the attached excel file.

Regards,

Field Manager"

Incentives

"Management has also decided to establish an incentive scheme based on the weekly sales performance of each Sales Rep. For each week, you can earn some extra money depending on how the sales of your stores compare to the goal:

- *If sales are below the goal figure, you will not earn a bonus*
- *If sales are above or equal to the goal figure, but below the stretched goal figure, you will earn a bonus of 10 dollars*
- *If sales are above or equal to the stretched goal figure, but below the super stretched goal figure, you will earn a bonus of 20 dollars*
- *If sales are above or equal to the super stretched goal figure, you will receive a bonus of 30 dollars*

If you worked in multiple stores within a single week, your bonus will be based on the ratio of the sum of your sales to the sum of the goal figures for those stores. In case you run sales events of both product A and product B in the same week, your bonus will be based on the average of the sales of those products weighted by the number of hours you worked for each. Your bonus in the end will be 0, 10, 20 or 30 dollar/week.

XXX from payroll will send you a weekly e-mail with the information about the bonus from the prior week.
The payment of all the bonuses will be made in June as a lump sum."

Feedback

"Management has also decided to provide you with weekly information regarding your sales performance by store and product-category. This information will show you, the goals and the actual sales, classified by store and product-category.

XXX from operations will e-mail you this data weekly."

Recognition

"Management has also decided to implement special acknowledgements for the Reps. The award categories are the following:

- *Gold Rep! Outstanding performance!*
  If sales are above or equal to the super stretched goal figure
- *Silver Rep! Awesome performance!*
  If sales are above or equal to the stretched goal figure, but below the super stretched goal figure
- *Bronze Rep! Good performance!*
  If sales are above or equal to the goal figure, but below the stretched goal figure
- *Promising Rep*
  If sales are below the goal figure

If you have multiple stores within a week, your category will be computed based on the sum of your sales compared to the sum of goals for those stores. In case you have events of product A and product B in the same week, your category will be an average of how you did in these products weighted by the number of hours worked in each.

In June you will receive a certificate with your Gold, Silver or Bronze Rep Recognition, according to the most prevalent category you qualified for throughout the campaign. Your name and category will also be announced by the company's newsletter.

Every week I'll e-mail you letting you know in what category you were in the prior week."

Appendix 2 – Example of the weekly incentives communication

"Hello,

Here is the update of your weekly bonus:
Your bonus for the week of May 17th-May 23rd is 30 [20/10/0] dollars.

Regards,

XXX
Payroll"

Appendix 3 – Example of the weekly feedback communication

"Hello,

Please find below the most recent information for the stores you worked in:
In the week of May 3rd-May 9th, store #X sales were 1 unit of product A (the Goal was 4, the Stretched Goal 5 and the Super Stretched Goal 6) and 6 units of product B (the Goal was 14, the Stretched Goal 15 and the Super Stretched Goal 16). Store #Y sales were 7 units of product B (the Goal was 13, the Stretched Goal 14 and the Super Stretched Goal 15).

Regards,

XXX
Operations"

Appendix 4 – Examples of the weekly recognition communication

"Hello,

Your award category for the special acknowledgement program is the following:
In the week of May 17th-May 23rd you were part of our

[Gold group of Reps. Congratulations! You are part of our most valuable group of Reps. Keep up the great work!]
[Silver group of Reps. Congratulations! We love to have you in our team. Keep up the good work!]
[Bronze group of Reps. Congratulations! You are part of our distinguished group of Reps. Keep it up!]
[Promising group of Reps. You are not yet part of our distinguished group of Reps. Give it a push and you can make it!]

Regards,

XXX
Field Manager"