MAN ON THE BENCH: THE POWER OF THE STICK IN THE SALES FORCE

MICHAEL J. AHEARNE
University of Houston

THOMAS J. STEENBURGH
University of Virginia

JEFFREY P. BOICHUK
University of Houston

FLORIAN KRAUS
University of Mannheim

CARMEN LIUTEC
University of Houston

August 2012

Michael J. Ahearne is C.T. Bauer Professor of Marketing and Executive Director, Sales Excellence Institute, University of Houston, C.T. Bauer College of Business, 334 Melcher Hall, Houston, TX 77204-6021 (Tel: 713.743.4155, Fax: 713.743.4572, Email: mahearne@uh.edu). Thomas J. Steenburgh is John L. Colley Associate Professor of Marketing, University of Virginia, Darden School of Business, 100 Darden Boulevard, Charlottesville, VA 22903 (Tel: 434.982.2392, Fax: 434.924.4724, Email: steenburght@darden.virginia.edu). Jeffrey P. Boichuk is a Marketing Doctoral Student, University of Houston, C.T. Bauer College of Business, 334 Melcher Hall, Houston, TX 77204-6021 (Tel: 713.743.4579, Fax: 713.743.4572, Email: jpboichuk@uh.edu). Florian Kraus is Dr. Werner Jackstädt Endowed Chair of Business Administration and Marketing IV, University of Mannheim, Schloss, 68161 Mannheim, Germany (Tel: +49.621.181.2662, Fax: +49.621.181.2672, Email: kraus@bwl.uni-mannheim.de). Carmen Liutec is a Marketing Doctoral Student, University of Houston, C.T. Bauer College of Business, 334 Melcher Hall, Houston, TX 77204-6021 (Tel: 713.743.4577, Fax: 713.743.4572, Email: cliutec@uh.edu).

Man on the Bench: The Power of the Stick in the Sales Force

Abstract

In order to manage and motivate salespeople, companies regularly use both carrots (e.g., commission) and sticks (e.g., punishment). Although sticks are commonly found in practice, astonishingly little academic research has been devoted to studying their effects. In this project, we study a novel program called the "man-on-the-bench" program in a field experiment setting. It was developed to manage low-performing salespeople and is similar to a forced-ranking system, wherein low performers are let go. In treatment districts, inactive "bench" salespeople were hired, and at the end of the program these salespeople took the place of the lowest performing salesperson in their districts if that salesperson did not make quota. In control districts, the company operated business as usual. We find the man-on-the-bench program to have an immediate and sustained impact on low-performing salespeople. Furthermore, the program had a greater impact in cohesive sales districts. This finding suggests that cohesive groups bonded together to fight off an outside threat, a theory that we subsequently support in a lab experiment. These results suggest that sticks can be used to effectively manage salespeople in the short run, and we are hopeful that this study opens the door to future work.

Key words: sticks, forced ranking systems, helping behavior, social cohesion, and salesperson performance.

---

1. Introduction

A broad range of analytic, empirical, and experimental research has studied sales force compensation plans (e.g., Basu et al. 1985; Coughlan and Narasimhan 1992; Lim et al. 2009; Steenburgh 2008). Primarily focused on the design of optimal incentive contracts, this research has generated various prescriptions for the use of compensation instruments known informally as "carrots" (e.g., variable pay, bonuses, contests). For instance, a general finding is that high-powered incentive contracts should be used when salesperson effort is costly to monitor (see Banerjee et al. 2012 for a recent review). However, missing from this research are prescriptions for the use of accountability instruments, known informally as "sticks" (e.g., threatened punishment, social pressure).

This omission exists despite estimates suggesting that somewhere between 25% and 50% of *Fortune 500* companies aggressively use threatened punishment to manage their bottom performers (Grote 2005). Threatened punishment is often levied with rules that companies devise for salesperson dismissal or with employee fines and penalties. Jack Welch implemented the most widely known example of this practice during his time as CEO of GE. During his tenure, the bottom 10% of performers were fired on an annual basis (Welch 2001). Other companies devise rules to threaten punishment based on performance floors. They mandate the attainment of certain commission levels or enforce minimum year-over-year growth objectives to determine firing decisions (Smith 2012). In a similar vein, some companies use stunning ways to single out poor performing salespeople by way of social pressure; examples include attaching balloons to poor performers' shoulders while they're in the office, constructing makeshift "penalty boxes" for them to sit in, and requiring them to sit on toilet seats at annual meetings.

Reasons for threatening punishment with top-down accountability systems or enacting social pressure programs stem from sales managers' hesitation toward enforcing negative incentives with low performing salespeople. In addition, when deciding whether or not to replace chronic low performers, sales managers generally place too much weight on sales force turnover costs, including costs associated with recruitment and selection, training, and territory vacancy (Darmon 1990). These costs are no doubt an unavoidable consequence of dismissing salespeople in many instances. But successful companies understand that removing or, at the very least, taking corrective action on poor performers is necessary to remain competitive.

With sticks playing such an important role in companies' talent management practices, it may seem puzzling that an established empirical literature has yet to develop. But as Scullen et al. (2005) state, there is "almost no supporting research base currently available" (p. 26). What is clear is that the primary reason for such scarcity is that data are difficult to obtain. Even the sales force incentive literature has suffered from a shortage of empirical data (cf. Misra and Nair 2011). Seeing as sticks are a more sensitive issue to companies than carrots are, researchers face a particularly uphill battle collecting data in this realm. As a result, no research has investigated the impact of sticks on salesperson performance based on company data. Aside from a stream of research based on experiments in the economics literature (e.g., Dickinson 2001; Fehr and Schmidt 2007), the extant literature is void of work in this area.

This paper exploits a field experiment to alleviate this shortcoming and, in doing so, studies a novel type of stick, which we term the "man-on-the-bench" program (hereafter, the MoB program). We use this nomenclature because the program relates closely to the impact that bench players have on starters in sports. It introduces inactive salespeople (i.e., bench players) into sales districts and imparts formal rules to determine whether they will replace current

salespeople (i.e., starters) after a pre-specified time period. In the application we studied, an inactive salesperson replaced the lowest performing salesperson in a district if the lowest performer failed to reach an annual target for the year the program was in place.

To this end, the MoB program overcomes two of the primary reasons why sticks are used in practice. First, it imposes a formal rule for dismissal to guard against sales managers' aversion to enforce corrective action. Second, it ensures that turnover costs, such as those associated with recruitment and selection, training, and territory vacancy, are decoupled from sales managers' dismissal decisions. Inactive salespeople are trained during the program such that they can be put into the starting lineup seamlessly.

In this paper, we address questions related to the impact of the MoB program (i.e., threatened punishment) on salesperson performance. Through our field experiment, we use a difference-in-differences estimation approach to find that the MoB program improved salesperson performance by approximately 4.12%. Furthermore, with regression analysis we find that threatening punishment affected the performance of bottom performing salespeople to a greater degree than it affected the performance of others. Evidently, social cohesion plays an important role, too. Greater performance was found in treatment districts that were more cohesive, and our results suggest that the helping behaviors of others may have contributed to bottom performers' increased performance. We investigate this phenomenon further in a controlled laboratory experiment, where participants could devote selling effort to both themselves and their group members. Indeed, when an external threat (i.e., a MoB) was present, helping behaviors were greater in cohesive groups than they were in non-cohesive groups, and these helping behaviors largely contributed to bottom performers' increased performance.

2. Background and Theory

Similar to traditional forced ranking systems, the MoB program is a talent management tool that is best applied in companies with performance-based cultures (see Table 1). Essentially, forced ranking systems weed out poor performers, making them most effective when "dead wood" still exists in an organization (Boyle 2001). Although Jack Welch implemented GE's forced ranking system throughout his tenure as CEO (Welch 2001), Scullen et al.'s (2005) simulation study suggests that forced ranking systems generate the greatest improvement for companies in the first two years. Furthermore, previous authors have suggested that an evidence-based management approach should be taken prior to the implementation of performance evaluation systems similar to the MoB program (Pfeffer and Sutton 2006). To ensure a culture fit, the MoB program and forced ranking systems in general should only be applied in organizations that have performance-based cultures (e.g., sales organizations).

----- Insert Table 1 about here -----

Incongruent characteristics. Table 1 also communicates areas where the MoB program deviates from traditional forced ranking systems. While it is commonly prescribed that forced ranking systems should include positive incentives that preferentially favor top performers (e.g., Grote 2005; Hazels and Sasse 2008), the MoB program implemented in our field experiment held employee bonuses constant. This feature is convenient as it allows us to estimate the impact of threatened punishment on salesperson performance, holding positive incentives constant. The threatened punishment imposed in the MoB program also differed slightly from what is common. Whereas traditional forced ranking systems use a relative rating system to determine which employees will be let go (e.g., the bottom 10% are fired), the MoB program used a rating system that was both relative and absolute (i.e., the lowest performing salesperson in a district would be

replaced *if* that salesperson failed to reach an annual target) (c.f. Duffy and Webber 1974). Finally, perhaps the most distinguishing characteristic of the MoB program is that the new hire selection process took place prior to (as opposed to after) salesperson dismissal. This departure ensured that the company was not left with vacant territories after salespeople were let go at the end of the program, a cost that is inherent to traditional forced ranking systems. Next we outline our theoretical predictions regarding the impact of the MoB program on salesperson performance.

2.1 Theoretical Predictions

A divide separates pundits in the popular press regarding the efficacy of forced ranking systems, with some contending that they enhance employee performance (e.g., Axelrod et al. 2002; Grote 2005; Welch 2001) and others arguing just the opposite (e.g., Boyle 2001; Gary 2001; Gladwell 2002). This debate spans many organizational contexts and many different types of employment, but no empirical evidence exists to clarify whether these programs work. Our study intends to take a step in this direction by providing empirical evidence with respect to the impact of the MoB program in a sales force setting. Our expectation, based on the very nature of stick-based incentives, was that threatening punishment through the MoB program would improve salesperson performance, with a relatively stronger increase coming from bottom performers.

With respect to the impact of forced ranking systems on employee dynamics, the existing literature contends that competition rather than cooperation ensues when negative incentives, such as threatened punishment, are introduced (Garcia and Tor 2007; Gary 2001; Pfeffer and Sutton 2006). However, the opposite may be true as well, and our contention was that social cohesion would play a major role in this process. Social cohesion has been defined as “the resistance of a group to disruptive forces” (Gross and Martin 1952, p. 553-54). Cohesive work

groups may, then, react to the MoB program by *cooperating* to fend off the out-group threat as opposed to *competing*, and performance may rise as a result. This prediction is further supported by the phenomena of parochialism, defined as "preferences for favoring the members of one's own social group" (Bernhard et al. 2006, p. 912). A general finding in this respect is that in-group members demonstrate a propensity to favor one another over out-group members (Choi and Bowles 2007; Yamagishi et al. 1999). Since in-group/out-group comparisons are likely to be more salient in cohesive work groups, we expected the MoB program to be more effective in cohesive work groups than in sparsely connected work groups.

We tested these theoretical predictions with two studies. Study 1 is a field experiment that took place within a Fortune 500 company's sales force. To the best of our knowledge, it is the first study in the literature that empirically documents the impact of a stick-based program on salesperson performance. Study 2 is a controlled laboratory experiment that replicates and extends our findings from Study 1. By experimentally manipulating the social cohesiveness of work groups, we lend stronger support to the occurrence of parochialism in this setting.

3. Study 1: Field Experiment

3.1. Institutional Details and Field Experiment Description

Our field experiment took place in a U.S. based Fortune 500 company. During the time of the study the company made over $6 billion/year in revenue in the B2B marketplace, but the executive team was concerned that a significant proportion of its sales force was composed of low performers. The company's Vice President of Sales suggested threatening punishment with the MoB program as a remedy to this problem. Although some executives were concerned that the program was antithetical to the company's culture, the executive team decided that it was

worth testing whether the program would work and allowed us to study the program through a quasi-experiment.

The MoB study was designed as follows. In the treatment districts, sales managers announced the program to their districts on the first workday in January of the program year. Existing salespeople were told that a new salesperson (a man on the bench) would be hired into their district later in the year. This man on the bench would take the job of the lowest performing salesperson in their district if that individual failed to make his or her annual quota. Alternatively, if every salesperson in a district achieved quota, everyone would keep their jobs and the man on the bench would be reassigned to a different district in the company that had an open territory. Men on the bench learned the job by training with their district managers and did not help or shadow any of existing salespeople during the training period. Their time "on the bench" was spent attaining certifications and engaging in highly intensive technical training. As is typical in technical sales positions, salespeople at this organization are required to install and maintain heavy equipment in addition to their regular selling activities, resulting in an extensive ramp-up period for new sales hires.

We collected three years of data, two years of data before the program was announced and one year while it was running, to assess the impact of the program on salesperson performance. Although we started working with the company before the program was announced, we were not able to randomly assign the men on the bench to sales districts due to practical considerations. The company, however, agreed not to specifically target sales districts with under-achieving salespeople when deciding which districts would be included in the treatment group. We will follow up on this point later in the results section, where we further discuss issues related to the quasi-experimental design of our study.

In the control districts, operations continued business as usual and no men on the bench were hired. Financial compensation did not vary across the treatment and control groups. All salespeople were compensated with a fixed salary of about $40,000 per year and an annual bonus, which was awarded for reaching an annual target, of $2,000 per year. Typical of many sales organizations, the sales force was organized geographically. Overall, 286 salespeople worked in the 58 sales districts in the treatment condition and 413 salespeople worked in the 82 sales districts in the control condition.

We also collected social network data from the company *before the MoB program had been announced to the sales force*. Of interest to us from a social network perspective was the density of sales districts' advice networks, a measure that captures social cohesion (Friedkin 2004). As outlined previously, we expected tight-knit sales districts to perform better when an out-group threat was introduced than otherwise comparable loose-knit sales districts. We collected social network data from the 699 salespeople in the 140 sales districts. Specifically, we asked salespeople to name all of the colleagues in their sales district to whom they go to for help or advice about work related matters. This nomination method is classified as a recall task, which is known to be a reliable method for collecting social network data in small groups (Marsden 2011).

The timing of events over the course of the study is summarized in Figure 1. The intervention, which amounted to the implementation of the MoB program in a subset of the company's sales districts, was announced in January of Year3. To measure its impact we collected salesperson-level performance data from the company's records during all three years, utilizing monthly performance data during Year2 and Year3 as pre- and post-test data, respectively. Annual Year1 data are used as a proxy measure for salespeople's selling ability.

----- Insert Figure 1 about here -----

These performance data indicate salespeople's dollar sales relative to an expected individual-based sales target, which was set for each salesperson's territory across the company by an external firm that specializes in sales compensation. Sales-to-target numbers are useful since targets account for differences in such factors as territory potential, number of customers, size of customers, and seasonality. Sales-to-target is considered by many to be the best outcome performance metric for sales force research (Churchill et al. 1985; Zoltners et al. 2006), with a long history of successful applications in the marketing and sales literatures (e.g., MacKenzie et al. 1993).

To recap, the following variables were collected: (1) *social cohesion*, calculated as the density of a sales district's advice network[1] and (2) *salesperson performance* and *salesperson ability*, both measured with objective sales-to-target data. In addition, we gathered information from company records on *salesperson tenure*, measured as the number of years that salespeople have spent with the company, and *district size*, being the number of salespeople assigned to a sales district. The descriptive statistics of these variables are reported in Table 2.

----- Insert Table 2 about here -----

We addressed three research questions in this field experiment: (1) Does threatening punishment by means of having a "man on the bench" improve sales force performance? (2) Do bottom performers react more intensely to threatened punishment than other employees? and (3) Does social cohesion play a role in sales districts' reactions to an out-group threat?

[1] Density is computed as the total number of directed (in-coming and out-going) advice ties within a sales district, divided by the maximum number of possible directed ties (i.e., n*(n-1)) within a sales district (cf. Hanneman and Riddle 2005; Wasserman and Faust 1994).

3.2. Results

Since our data come from a quasi-experiment, our first concern was to test whether the treatment and control groups contain comparable observations. The initial balance between the treatment and control groups is reported in Table 3. The standardized mean difference, a commonly used measure to check balance, is the difference in means between the treatment and control groups divided by the standard deviation of the treatment group. A rule of thumb is that better balance between the groups is needed if the standardized mean difference is greater than 0.25 (Ho et al. 2007, p. 221). As can be seen in Column 3 of Table 3, the treatment and control groups are reasonably balanced across the variables that conceivably could have entered into the firm's decision about treatment allocation. In particular, notice that the two variables measuring salespeople's past performance, Sales-to-Target Yr1 and Sales-to-Target Yr2, fall well below the cutoff of 0.25. This confirms that the company, as it claimed, did not systematically assign men on the bench in greater proportions to underperforming districts.

----- Insert Table 3 about here -----

The treatment and control groups, however, were out of balance on the social cohesion variable. A standardized mean difference of 0.48 indicates that the treatment group was composed of sales districts with more densely knit social networks than was the control. Accordingly, we used a variety of matching procedures to bring these two groups into an acceptable range of balance on this variable. Although matching techniques are not yet commonly used in marketing, they have gained wide acceptance in other fields in the social

sciences.[2] We used the matching strategy and software suggested by Ho et al. (2012) to balance the groups because it includes a broad, yet reasonable number of popular techniques.

For our data, a genetic matching algorithm (Diamond and Sekhon 2005) produced the best match between the treatment and control groups. Unlike the genetic matching algorithm, many of the other procedures we applied, such as exact and optimal matching, failed to achieve acceptable levels of balance on social cohesion. With the genetic matching algorithm, the standardized mean difference between the groups on the social cohesion variable fell from 0.48 in the unmatched data to 0.02 in the matched data (see Column 6 of Table 3). Furthermore, all other control variables continued to show reasonable degrees of balance in the matched data. The greatest standardized mean difference across all variables is less than 0.05, which falls well below the 0.25 cutoff. The change in balance is graphically depicted in Figure 2. The distance variable represents the change in propensity scores. The matched dataset was attained without sacrificing any of the treatment observations, but did eliminate 249 of the control observations. This leaves us with 286 salespeople in the treatment condition and 164 salespeople in the control condition.

----- Insert Figure 2 about here -----

Having attained comparable treatment and control groups, we proceeded to test the effectiveness of the MoB program through difference-in-differences estimation. We present this analysis in Table 4. This calculation suggests that the MoB program had a positive and significant impact on salespeople's performance (effect magnitude of 4.12, $p < .001$). On average, salespeople in the treatment districts performed 4.12 percentage points higher than would be expected had they been in control districts. This positive treatment effect is graphically

[2] In marketing, an example is Avery et al. (2012). In other fields, examples include Winship and Morgan (1999) in sociology; Lee and Wahal (2004) in finance; Ho et al. (2007) in political science; Jaffe et al. (1993), Meyer (1995), and Heckman et al. (1998) in economics; and Hansen (2004) in education.

depicted in Figure 3. As can be seen in the figure, salespeople's performance improved in the first month that the program was put into place and continued at this higher level throughout the year. By comparison, the control group's performance did not change across the two years.

----- Insert Table 4 and Figure 3 about here -----

This difference-in-differences estimate can easily be translated into a regression framework that allows for a more detailed analysis. Specifically, we wanted to understand how the effects of the program varied across high and low performers and across groups with different levels of social cohesion. We specified the model:

$$y_{it} = \alpha + \sum_{m=1}^{11} \tau_m M_{mt} + \tau_{12} Tenure_i + \tau_{13} Size_k + \delta_1 A_i + \delta_2 C_k + \delta_3 Yr3_t + \delta_4 T_{it} + \delta_5 MB_{it} + \delta_6 MB_{it} * A_i + \delta_7 MB_{it} * C_k + \epsilon_{it}$$

where:

y_{it} = performance (sales-to-target) of salesperson i during time t (months = 13, …, 36);

M_{mt} = calendar month dummy variables (February - December);

$Tenure_i$ = company tenure for salesperson i (in years);

$Size_k$ = the number of salespeople in sales district k.

A_i = selling ability of salesperson i (sales-to-target in Year1);

C_k = cohesion in sales district k (advice network density);

$Yr3_t$ = a dummy variable equal to one in Year3, zero in Year2;

T_i = a dummy variable equal to one if salesperson i is in the treatment condition (the MoB program), zero otherwise;

MB_{it} = a dummy variable equal to one in Year3 if salesperson i is in the treatment condition (the MoB program), zero otherwise;

We present the regression results in Table 5. The estimation of the model accounts for correlation across time for the same individual, with robust standard errors clustered by salesperson.[3]

----- Insert Table 5 about here -----

Consistent with our difference-in-difference estimate, Model 1 in Table 5 shows that the MoB program increased the average performance of salespeople by about four percentage points ($\delta_5 = 4.124$, $p < 0.001$). Model 2 expands on this analysis by showing that the MoB program had a greater impact on low ability salespeople. To be specific, we found a significant negative interaction between the MoB program and selling ability ($\delta_6 = -0.399$, $p < 0.001$). We graphically depict this effect in Figures 4 (control condition) and 5 (MoB condition). Both figures illustrate salesperson performance (i.e., sales-to-target) along a time series from Month 13 (January of Year2) – Month 36 (December of Year3) by selling ability quartiles (i.e., Year1 sales-to-target quartiles). In Figure 5, we see that low ability salespeople immediately increased their performance by about eight percentage points when the program was announced and sustained this level throughout the year the program was run. In contrast, the performance of high ability salespeople remained unaffected by the MoB program. This depiction is in line with our expectation that poor performing salespeople would be the most responsive group to threatened punishment. In Figure 4, we show that the performance of each segment of salespeople in the control districts did not change across the two years.

----- Insert Figures 4 and 5 about here -----

Model 2 in Table 5 also explores how the effect of the MoB program differed across varying levels of social cohesion. The interaction term between social cohesion and the MoB program is positive and significant ($\delta_7 = 11.143$, $p < 0.001$), indicating that the program effect

[3] Clustering by salesperson and district yielded similar results.

was even higher in cohesive sales districts. The results of a post-hoc test, which we report in the following section, suggest that top performers in cohesive sales districts may have engaged in helping behaviors at the end of the MoB program, which boosted the performance of these districts.

3.2. Discussion

Our field study provides compelling empirical evidence in favor of the use of sticks as motivational tools. The immediate effect of the MoB program was a boost in performance by approximately four percentage points. Furthermore, the program had the desired effect of pushing the lowest performing salespeople to improve performance. The immediate jump in performance from this group suggests that low performers may have been shirking prior to the introduction of the program.

We also found that the MoB program was more effective when it was implemented in tightly knit sales districts. We conjecture that cohesive sales districts could have performed better as a result of two processes. First, salespeople who were sure to hit their target may have helped their low performing colleagues, thereby favoring an in-group member over the man on the bench. This is intriguing because the looming threat of the program may have heightened the goodwill among salespeople in tight-knit districts and caused them to band together. Alternatively, salespeople in tight-knit districts may have been more motivated to keep their jobs because they enjoy working with their group and increased their selling effort to hit quota. These processes offer two mechanisms through which performance in cohesive sales districts could have developed: helping effort and own effort.

We investigated the possibility of helping effort with a post-hoc test of the correlation between the performance of the highest and lowest performers in each sales district in the last six

months of the program by condition.[4] Only in the MoB – High Cohesion condition was the correlation negative and significant (-0.13, $p < .05$). In all other conditions, the correlation was either non-significant or positive and significant. Based on these results and post-program interviews with salespeople and managers involved with the program, we speculate that, within cohesive sales districts, help from salespeople in the first quartile (top performers) may have partially contributed to the incrementally better performance of salespeople in the fourth quartile (bottom performers). If so, these results interestingly lend support to the notion of parochialism (in-group favoritism), a possibility that we investigate further in a controlled setting in Study 2.

4. Study 2: Laboratory Experiment

4.1. Objectives

Next, we directly investigate salesperson effort and performance by means of a laboratory experiment. We ran this experiment with two objectives and one inquiry in mind. Our two objectives were (a) to replicate the finding that adding a MoB increases performance of salespeople in a controlled laboratory setting and (b) to further validate the finding that the MoB effect is stronger when group cohesion is high by experimentally manipulating social cohesion. Our inquiry and primary extension beyond Study 1 relates to our second objective. Specifically, we aimed to understand a possible mechanism through which social cohesion moderates the MoB effect: helping behavior. From our field study, we posited that low performing salespeople in cohesive sales districts performed better under the MoB program in part due to helping behaviors from higher performing salespeople. A controlled laboratory setting allows us to investigate this possibility because it enables us to measure how individuals allocate effort to

[4] For this post-hoc test, high and low cohesion conditions were achieved based on a median split of our social cohesion variable.

themselves and to others. We expect both performance and helping behaviors to be greatest when a MoB is present and group cohesion is high.

4.2. Laboratory Experiment Description

Each participant was assigned to a group of four people (groups can be thought of as sales districts). They were asked to make eight effort allocation decisions over two periods, with four decision rounds per period (decisions rounds are analogous to quarters and periods are analogous to years). Effort could be applied to one's own task, to a fellow group member's task (i.e., helping behavior), or to both. Participants received a salary of $3 per period and an additional bonus of $5 if they reached an individual-based target, which we set at 400 points per period. Decision numbers for one's own task, as well as helping behavior decision numbers, were bounded between [0, 100]. Both types of effort had identical quadratic cost functions $C(Effort) = k*Effort^2$ associated with them, with $k = 0.0002$. Unobserved demand factors were incorporated with a random term that was uniformly distributed over the range [0, 80]. A participant's total amount of points after each decision round amounted to the sum of three parts: (a) one's own effort decision number, (b) the helping decision points received from fellow group members, and (c) a random term. Total points for decision rounds one to four were accumulated and contributed to participants' target attainment. A detailed description of the experiment's expected payouts can be found in Appendix A.

The experiment amounts to a 2 (treatment: control, MoB) × 2 (social cohesion: low, high) factorial design with random assignment. At the beginning of the experiment, participants were either told that everyone would engage in the decision rounds for both periods regardless of performance (control condition) or that a MoB would replace the lowest performing participant

who missed target after the first period (MoB condition).[5] In addition, prior to the experiment, participants either introduced themselves to the *room* (comprised of five groups) and participated in a word problem exercise with everyone in the *room* (low group cohesion condition) or introduced themselves to their *group members* and participated in a word problem exercise with their *group members* (high group cohesion condition).

All participants were given detailed instructions along with cost of effort numbers in tabular form on a sheet of paper. They also engaged in one practice round to gain familiarity with the computer interface. We recruited 20 participants from a public university in the Southern United States for each condition. Fifty-five percent of these participants were female, their average age was approximately 23 years old, and the average participant had 2.16 years of selling experience.

4.3. Manipulation Check and Results

After the first period, participants completed a questionnaire that included an established scale for group cohesion (Carless and De Paola 2000; Chin et al. 1999) (see Appendix B). A manipulation check shows that participants in the high cohesion condition rated group cohesion significantly higher than participants in the low cohesion condition (5.68 vs. 3.87, $t = 4.84$, $p < .001$). Accordingly, these results lend support to the effectiveness of our manipulation.

Figure 6 depicts participants' total sales (which includes own effort, help received, and a random term) for the decision rounds in period one. Figures 7 and 8 decompose this graph into two of its parts, those being own effort and helping effort, respectively. The decision rounds in period one are of interest because they represent the time when the MoB program was in place for the treatment conditions. Thus, our analyses will focus on total sales, own effort, and helping

[5] We actually had an additional participant per group literally "sitting on the bench," waiting to replace the lowest performer who missed target after the first period.

effort at the aggregate level for the first period.

----- Insert Figures 6, 7, and 8 about here -----

Figure 6 shows that participants' performance was similar up until the final round, when the performance of participants in the MoB – High Cohesion condition exceeded that of the other conditions. A t-test reveals that participants in the MoB condition performed significantly better than participants in the control condition in period one (390.70 vs. 360.33, $t = 2.41$, $p = 0.01$). Thus, threatened punishment by means of the MoB program led to higher performance in the laboratory, replicating our findings from Study 1. Furthermore, the MoB program had a greater impact in cohesive groups. In line with our expectations, results show that the interactive effect of treatment and social cohesion is positive and statistically significant ($F_{1, 79} = 14.14$, $p < .001$).

Interestingly, our experimental design allows us to understand what type of effort (own effort or helping effort) contributed to the success of the MoB program in cohesive groups. Figure 7 illustrates that participants expended similar levels of effort toward themselves regardless of their experimental condition assignment. A 2 (treatment) x 2 (cohesion) ANOVA with total own effort (i.e., own effort aggregated at the period level) as the dependent variable supports this visual representation. Neither the main effects nor their interaction effect were significant (Fs < 1.90). By contrast, Figure 8 illustrates that participants expended greater levels of helping effort when a MoB was present and group cohesion was high. Supporting the notion that the improved performance in the MoB – High Cohesion groups came from helping behaviors, the interactive effect of treatment and social cohesion on helping effort is statistically significant and positive ($F_{1, 79} = 6.11$, $p < .05$). This finding lends direct support to the notion of parochialism in the presence of an out-group threat. It also provides a strong test to our theory regarding helping behaviors as a possible mechanism through which social cohesion moderated

the impact of the MoB program on performance in Study 1.

5. General Discussion

Although the extant sales literature has established prescriptions for the use of carrots in sales compensation plans, empirical work in the area of sticks is still nascent. We exploit a rich dataset from a Fortune 500 company's sales force that includes objective pretest and posttest performance data as well as social network data to address this gap in the literature. We also subjected the empirical findings to a controlled experiment setting to replicate and extend the results from the field. Taken together, these two studies offer consistent evidence that threatened punishment has a positive impact during the implementation of a formal stick program. Threatening punishment enhanced salesperson performance in both the field and the lab. Further, this impact was most pronounced for poor performers and tight-knit sales districts, with the increased performance in cohesive districts stemming from helping behaviors.

5.1. Theoretical Contributions

This paper contributes to theory development not only in the ongoing debate about forced ranking systems, but also in the sales and group dynamics literatures. First, the findings we report herein should be well received by participants in the ongoing debate regarding the efficacy of forced ranking systems (e.g., Axelrod et al. 2002; Boyle 2001; Gary 2001; Gladwell 2002; Grote 2005; Welch 2001). At least in a sales force setting, our results lend credence to the positive impact that forced ranking systems have on performance. Future research conducted in workplace contexts outside of the sales force setting will help delineate the external validity of these results.

Second, we add to a growing body of literature in the sales domain that suggests

salespeople have differential responses to incentives, dependent upon their position along the performance curve (e.g., Chung et al. 2012). We see that threatened punishment has a significant performance impact on low performance segments and no impact on high performance segments. The sales discipline would be advanced considerably if more research addressed salespeople's heterogeneous responses to incentives.

Third, our result pertaining to the contingency role that social cohesion plays contributes to research in the area of in-group favoritism (e.g., Yamagishi et al. 1999), and suggests that social cohesion plays an important role in a group's reaction to an external threat. Salespeople and lab participants in cohesive groups effectively made the decision to support their in-group members rather than to allow an out-group threat to enter their group.

5.2. Managerial Contributions

Many practitioners have argued that forced ranking systems create a culture of accountability that leads to positive performance outcomes (e.g., Axelrod et al. 2002; Grote 2005; Welch 2001). Our study provides empirical evidence to support this claim and may help explain the widespread use of forced ranking systems in practice. As soon as the MoB program was implemented, the threat of losing their job prodded low-performing salespeople to work harder and improve their performance. Furthermore, salespeople at this company did not devolve into a ruthless competition for resources, as others suggest is often the case (e.g., Boyle 2001; Gary 2001; Gladwell 2002). Rather, our data suggest the opposite; salespeople in tight-knit districts banded together to fight off the outside threat. Managers who need to make cuts to their sales force, whether these cuts are due to a high proportion of shirkers or to eroding business conditions, might consider the MoB program as a possible solution.

While our results suggest that forced ranking systems can be effective, managers should

be cautious not to over-interpret our findings. Although the MoB program was effective in the short run, the company did not have a performance-based culture and chose to abandon the program rather than to extend it to other sales districts. Given this, we do not know whether the program would have remained effective if it became a standard policy. For instance, we worry that the program may have eroded the social cohesion of the tight-knit sales districts and lost its effectiveness or become counter-productive over time if it became a standard policy.

5.3 Future Research Directions

Research avenues we view to be fruitful directions for future work in the area of sticks pertain to the impact such programs have on social cohesion and on long-term performance. First, it would be interesting to know not only how the MoB program affects sales districts' performance over time, but also how it affects sales districts' social cohesion over time. Again, it may very well be the case that the MoB program, if implemented for a longer period, would damage districts' social cohesion, reducing the overall effectiveness of the program. There is some support for this idea in the popular management literature, as some suggest that forced ranking systems foster a competitive climate (Garcia and Tor 2007; Gary 2001; Pfeffer and Sutton 2006). At the current time, anecdotal evidence is the best source of insight we can provide.

At the company, the Vice President of Sales considered the program to be overwhelmingly successful. Not all of the salespeople who stayed with the company viewed the program as being fair, but frontline sales managers suggested that many of the average and star performers thought it was an equitable way to make cuts. The program was based on attainable and quantitative metrics, making it an objective way to make sales force turnover decisions. Therefore, anecdotally we can say that the program was deemed to be fair by many, but an empirical study in this thrust would be a welcome extension to our paper.

Second, it is important to note that our results pertain to the impact of threatened punishment in the short term and, more specifically, its impact during the time that punishment was threatened as opposed to realized. In the name of future research, an additional mention of the lasting effect of the MoB program is in order. With 47 salespeople being replaced at the end of Year3 in our field experiment, the organization endured a shock to its culture, which had been established based on a long history of supporting its employees. This shock ultimately resulted in the MoB program being discontinued after its first year in operation. But the primary reason, again, was due to cultural concerns, not the short-term impact of threatening punishment, which was the focus of this paper. This outcome further emphasizes the importance of adopting an evidence-based management approach before enacting stick-based programs. Successful programs most often stem from bottom-up processes that are supported by the members they ultimately impact.

Studying the long-term effects of sticks fell outside the scope of this paper, but we offer this extension as an interesting avenue for future research as well. For instance, understanding the conditions under which the short-term performance increases resulting from threatened punishment extend into long-term performance increases would be of practical interest to managers and of theoretical interest to academics.

References

Avery, J., T.J. Steenburgh, J. Deighton, M. Caravella. 2012. Adding bricks to clicks: Predicting the patterns of cross-channel elasticities over time. *Journal of Marketing* **76**(3) 96-111.

Axelrod, B., H. Handfield-Jones, E. Michaels. 2002. A new game plan for c players. *Harvard Business Review* **80**(1) 80-88.

Banerjee, R., M. Bergen, S. Dutta, S. Ray. 2012. Applications of agency theory in B2B marketing: Review and future directions. G. Lilien, R. Grewal, eds. *Handbook of business-to-business marketing*. Edward Elgar, Northampton, MA, USA, 41-53.

Basu, A.K., R. Lal, V. Srinivasan, R. Staelin. 1985. Salesforce compensation plans: An agency theoretic perspective. *Marketing Science* **4**(4) 267-291.

Bernhard, H., U. Fischbacher, E. Fehr. 2006. Parochial altruism in humans. *Nature* **442**(7105) 912-915.

Boyle, M. (2001, May 28). Performance reviews: Perilous curves ahead. Fortune. **143**(11) 187-188.

Carless, S.A., C. De Paola. 2000. The measurement of cohesion in work teams. *Small Group Research* **31**(1) 71-88.

Chin, W.W., D. Salisbury, A.W. Pearson, M.J. Stollak. 1999. Perceived cohesion in small groups: Adapting and testing the perceived cohesion scale in a small-group setting. *Small Group Research* **30**(6) 751-766.

Choi, J.K., S. Bowles. 2007. The coevolution of parochial altruism and war. *Science* **318**(5850) 636-640.

Chung, D., T.J. Steenburgh, K. Sudhir. 2012. Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. Working Paper, Harvard Business School, Cambridge, MA.

Churchill, G.A., Jr, N.M. Ford, S.W. Hartley, O.C. Walker Jr. 1985. The determinants of salesperson performance: A meta-analysis. *Journal of Marketing Research* **22**(2) 103-118.

Coughlan, A.T., C. Narasimhan. 1992. An empirical analysis of sales-force compensation plans. *Journal of Business* **65**(1) 93-121.

Darmon, R.Y. 1990. Identifying sources of turnover costs: A segmental approach. *Journal of Marketing* **54**(2) 46-56.

Diamond, A., J. Sekhon. 2005. *Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies.*

Dickinson, D.I. 2001. The carrot vs. the stick in work team motivation. *Experimental Economics* **4**(1) 107-124.

Duffy, K.E., R.E. Webber. 1974. On "relative" rating systems. *Personnel Psychology* **27**(2) 307-311.

Fehr, E., K.M. Schmidt. 2007. Adding a stick to the carrot? The interaction of bonuses and fines. *American Economic Review* **97**(2) 177-181.

Friedkin, N.E. 2004. Social cohesion. *Annual Review of Sociology* **30**(1) 409-425.

Garcia, S.M., A. Tor. 2007. Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes* **102**(1) 95-108.

Gary, L. 2001. The controversial practice of forced ranking. *Harvard Management Update* **6**(10) 3-4.

Gladwell, M. (2002, July 22). The talent myth. The New Yorker. 28-33.

Gross, N., W.E. Martin. 1952. On group cohesiveness. *American Journal of Sociology* **57**(6) 546-564.

Grote, D. 2005. *Forced ranking: Making performance management work*. Harvard Business Review Press, Cambridge, MA.

Hanneman, R.A., M. Riddle. 2005. *Introduction to social network methods*. University of California, Riverside, Riverside, CA.

Hansen, B.B. 2004. Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association* **99**(467) 609-618.

Hazels, B., C.M. Sasse. 2008. Forced ranking: A review. *SAM Advanced Management Journal* **73**(2) 35-39.

Heckman, J., H. Ichimura, J. Smith, P. Todd. 1998. Characterizing selection bias using experimental data. *Econometrica* **66**(5) 1017-1098.

Ho, D., K. Imai, G. King, E.A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* **15**(3) 199-236.

Jaffe, A.B., M. Trajtenberg, R. Henderson. 1993. Geographic localization of knowledge spillovers as evidence by patent citations. *Quarterly Journal of Economics* **108**(3) 577-598.

Lee, P.M., S. Wahal. 2004. Grandstanding, certification and the underpricing of venture capital backed IPOs. *Journal of Financial Economics* **73**(2) 375-407.

Lim, N., M. Ahearne, S. Ham. 2009. Designing sales contests: Does the prize structure matter? *Journal of Marketing Research* **46**(3) 325-345.

MacKenzie, S.B., P.M. Podsakoff, R. Fetter. 1993. The impact of organizational citizenship behavior on evaluations of salesperson performance. *Journal of Marketing* **57**(1) 70-80.

Marsden, P.V. 2011. Survey methods for network data. J. Scott, P. J. Carrington, eds. *The sage handbook of social network analysis*. Thousand Oaks, California, 370-388.

Meyer, B.D. 1995. Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics* **13**(2) 151-161.

Misra, S., H.S. Nair. 2011. A structural model of salesforce compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics* **9**(3) 211-257.

Pfeffer, J., R.I. Sutton. 2006. Evidence-based management. *Harvard Business Review* **84**(1) 62-74.

Scullen, S.E., P.K. Bergey, L. Aiman-Smith. 2005. Forced distribution rating systems and the improvement of workforce potential: A baseline simulation. *Personnel Psychology* **58**(1) 1-31.

Smith, B. 2012. "When to say, you're fired!" GALLUP Business Journal. Retrieved on May 8, 2012 from: http://businessjournal.gallup.com/content19660/when-to-say-youre-fired.aspx.

Steenburgh, T. 2008. Effort or timing: The effect of lump-sum bonuses. *Quantitative Marketing and Economics* **6**(3) 235-256.

Wasserman, S., K. Faust. 1994. *Social network analysis: Methods and applications*. Cambridge University Press, New York, NY.

Welch, J. 2001. *Jack: Straight from the gut*. Warner Books, New York, NY.

Winship, C., S.L. Morgan. 1999. The estimation of causal effects from observational data. *Annual Review of Sociology* **25** 659-706.

Yamagishi, T., N. Jin, T. Kiyonari. 1999. Bounded generalized reciprocity: In-group boasting and in-group favoritism. *Advances in Group Processes* **16** 161–197.

Zoltners, A.A., P. Sinha, S.E. Lorimer. 2006. Match your sales force structure to your business life cycle. *Harvard Business Review* **84**(7/8) 81-89.

TABLE 1
A comparison of the man-on-the-bench program to traditional forced ranking systems

Relevant Characteristics	Traditional Forced Ranking Systems	Man-on-the-Bench Program
Congruent Characteristics		
Temporal Focus	Short-term	Short-term
Talent Management Tool	Yes	Yes
Organizational Culture Needed	Performance-based	Performance-based
Incongruent Characteristics		
Positive Incentives	Preferential	Held constant
Negative Incentives	Relative	Relative and Absolute
New Hire Selection Process	After dismissal	Prior to dismissal

TABLE 2
Descriptive statistics

	Control	MoB	Overall
Sales-to-Target Yr1	99.308	99.630	99.439
	(9.794)	(9.860)	(9.815)
Sales-to-Target Yr2	96.369	96.454	96.404
	(8.160)	(7.991)	(8.086)
Sales-to-Target Yr3	96.727	101.386	98.633
	(11.176)	(8.211)	(10.320)
Salesperson Tenure	9.908	9.507	9.744
	(6.514)	(6.200)	(6.386)
District Size	5.320	5.154	5.251
	(1.333)	(1.192)	(1.279)
Social Cohesion	0.118	0.184	0.145
	(0.113)	(0.138)	(0.128)
Number of Salespeople	413	286	699
Number of Sales Districts	82	58	140
Number of Salespeople Replaced	0	47	47

Notes: Standard deviations in parentheses.

TABLE 3
Treatment and control group comparison: Standardized mean differences

	All Data			Matched Data		
	Treatment Group Mean	Control Group Mean	Standardized Mean Difference	Treatment Group Mean	Control Group Mean	Standardized Mean Difference
Sales-to-Target Yr1	99.630	99.308	0.033	99.630	99.705	-0.008
Sales-to-Target Yr2	96.454	96.370	0.011	96.454	96.617	-0.020
Salesperson Tenure	9.507	9.909	-0.065	9.507	9.198	0.050
District Size	5.154	5.320	-0.139	5.154	5.112	0.035
Social Cohesion	0.184	0.118	0.477	0.184	0.181	0.022

TABLE 4
Difference-in-differences t-test for the matched data

	Sales-to-Target Yr2	Sales-to-Target Yr3	Difference (Yr3-Yr2)
MoB	96.454	101.386	4.932***
	(0.096)	(0.099)	(0.100)
Control	96.903	97.711	0.808
	(0.127)	(0.174)	(0.115)
Difference	-0.450	3.674***	4.124***
(MoB-Control)	(0.159)	(0.200)	(0.428)

Notes: Standard errors in parentheses; unequal variances assumed; the number of salespeople is 450.
*** $p < 0.001$.

TABLE 5
Regression results for matched data

	Model 1	Model 2
Intercept	98.716***	98.567***
	(0.670)	(0.644)
Covariates		
Calendar Month Dummies	Yes	Yes
Salesperson Tenure	0.010	0.011
	(0.028)	(0.026)
District Size	0.153	0.141
	(0.149)	(0.139)
Main Effects		
Selling Ability	0.688***	0.820***
	(0.020)	(0.017)
Social Cohesion	4.482**	0.142
	(1.488)	(1.201)
Year3 Dummy	-1.061	-1.061
	(0.836)	(0.836)
Treatment Dummy	-0.362	-0.128
	(0.241)	(0.206)
MoB Effect	4.124***	3.895***
	(0.750)	(0.715)
MoB Effect * Selling Ability		-0.399***
		(0.045)
MoB Effect * Social Cohesion		11.144***
		(3.378)
F-value	79.628***	135.824***
R^2	0.410	0.444
Model *F* change		13.763***

Notes: The dependent variable is monthly sales performance relative to quota for months 13 to 36; the number of observations is 10,800 and the number of salespeople is 450 (286 from treatment districts and 164 from control districts); standard errors are in parentheses; standard errors are clustered at the salesperson level.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

FIGURE 1
Field experiment timeline of events

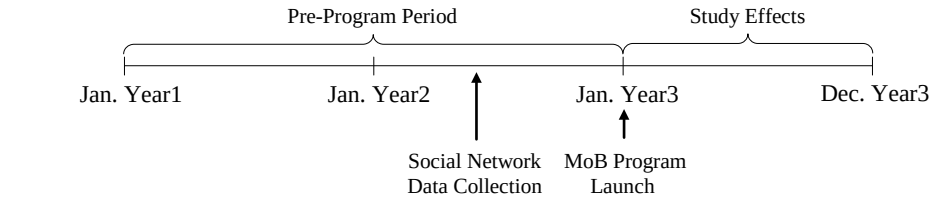


FIGURE 2
A graphical depiction of the change in balance before and after matching

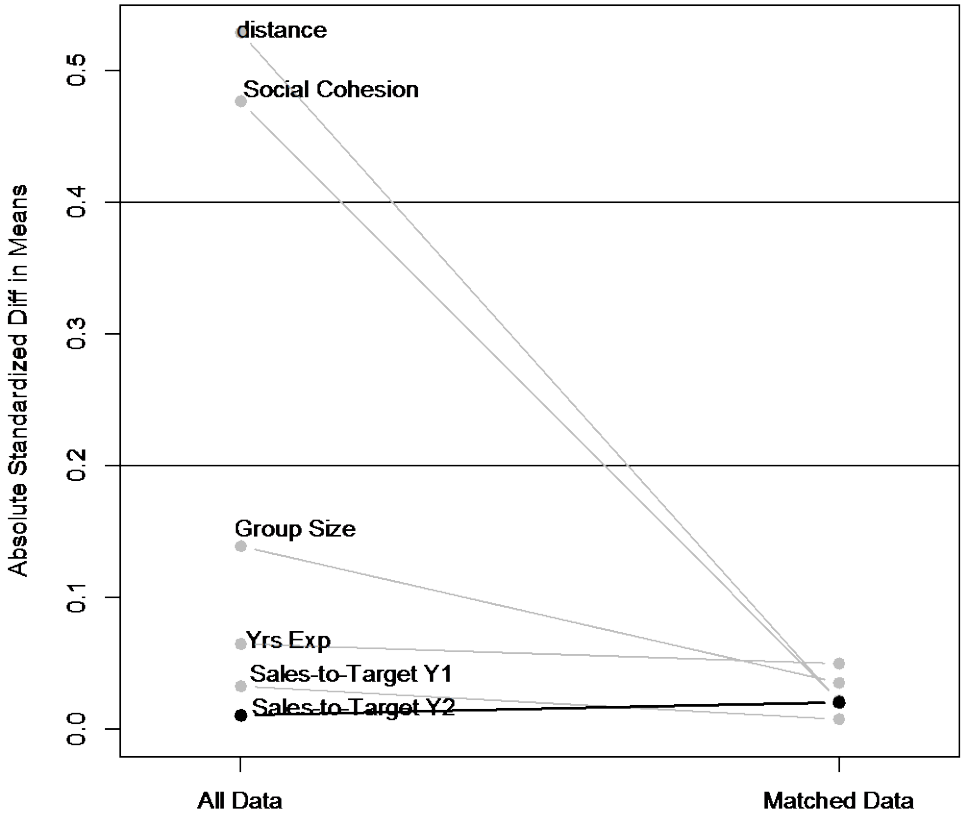


FIGURE 3
Monthly sales-to-target for Year2 and Year3 by condition

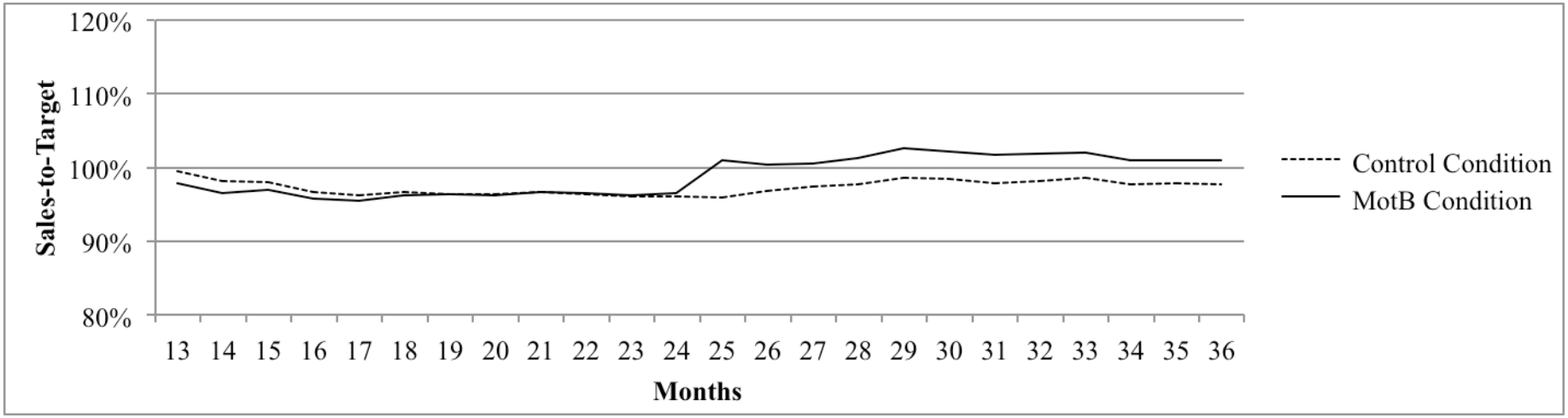


FIGURE 4
Control condition: Monthly sales-to-target for Year2 and Year3 by selling ability quartiles (Year1 sales-to-target quartiles)

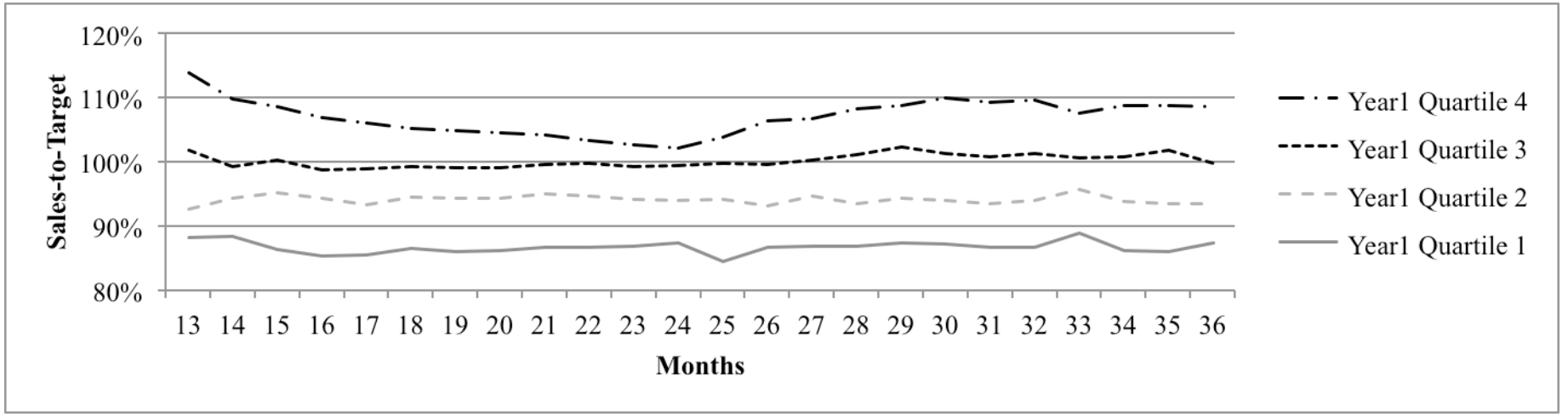


FIGURE 5
MoB condition: Monthly sales-to-target for Year2 and Year3 by selling ability quartiles (Year1 sales-to-target quartiles)

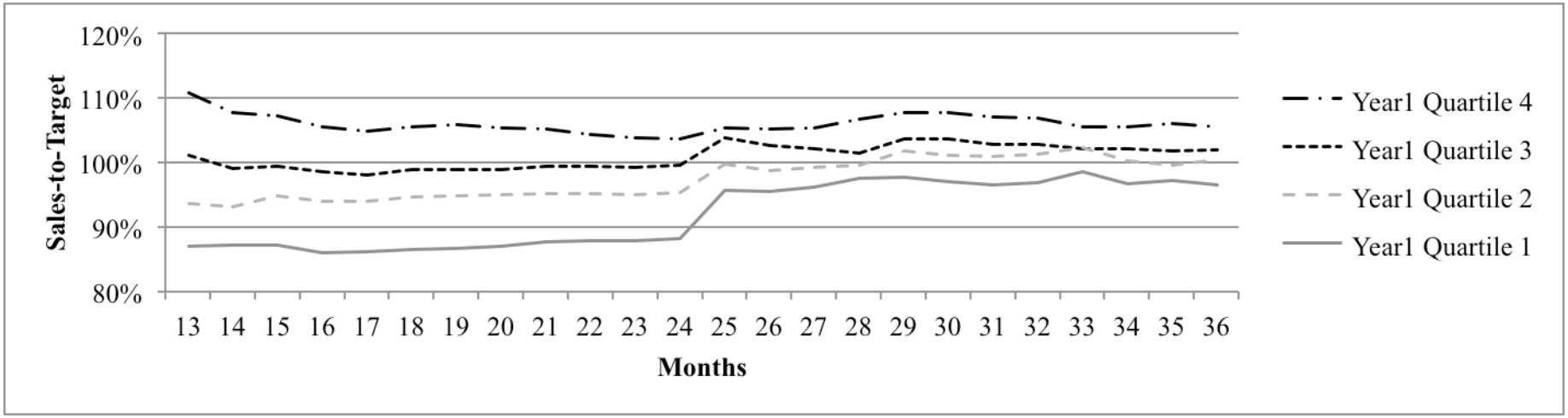


FIGURE 6
Total sales by decision round for each condition

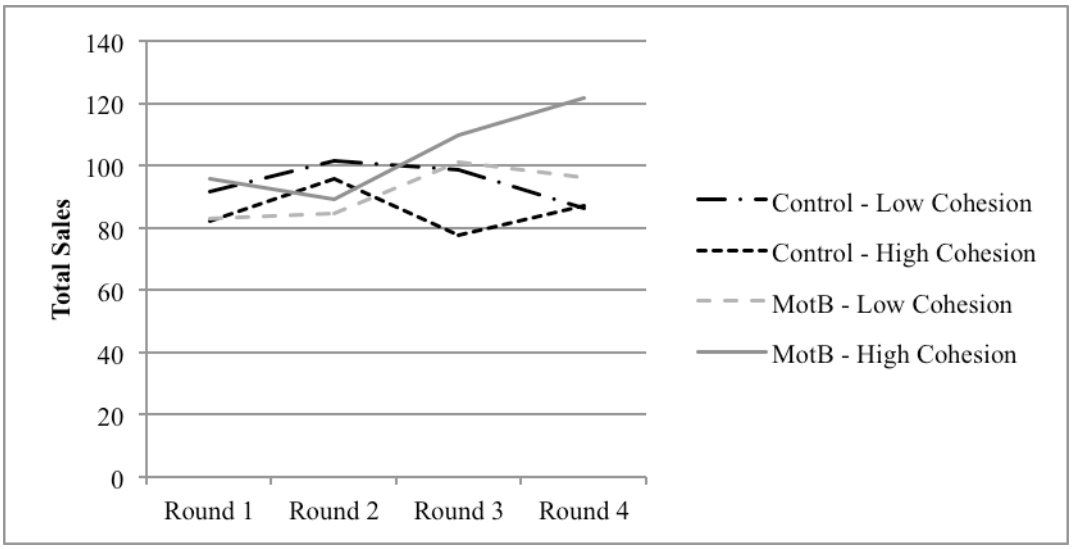


FIGURE 7
Own effort by decision round for each condition

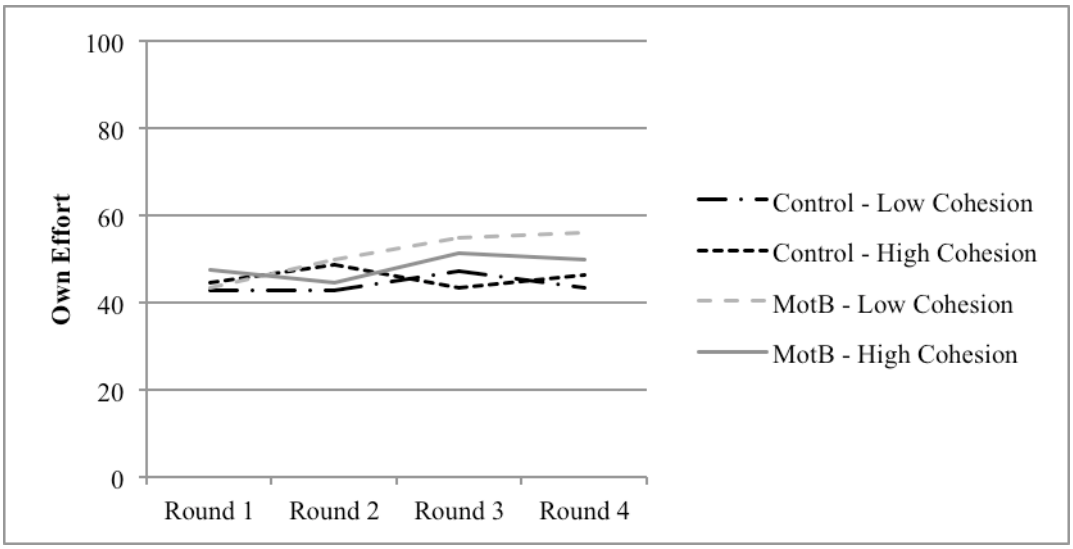
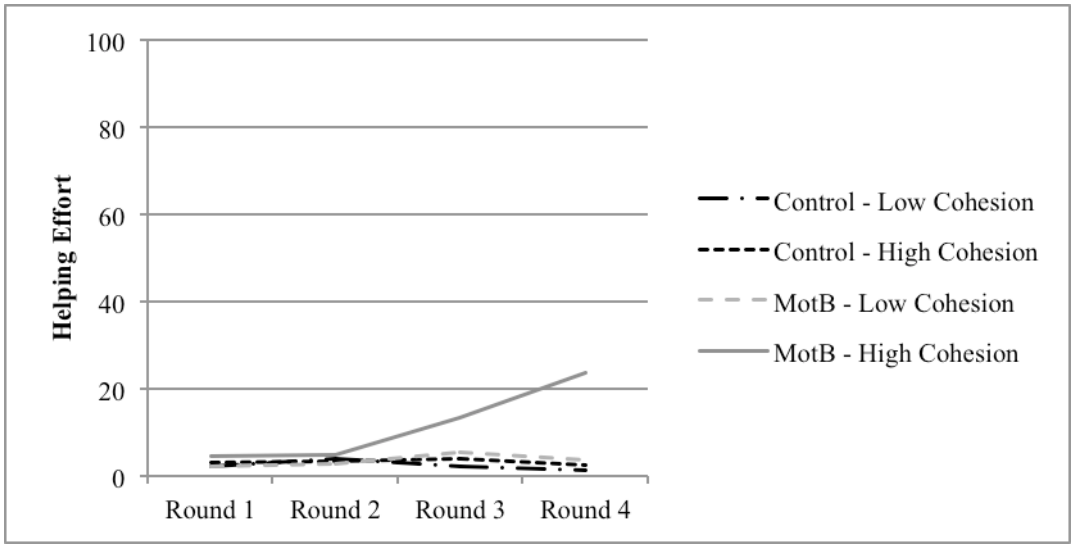


FIGURE 8
Helping effort by decision round for each condition



APPENDIX A
Laboratory experiment expected payouts (net of cost of effort)

	Control	**MoB**
Participation payout	$3.00	$3.00
Target bonus	$5.00	$5.00
Probability of reaching target each round, with effort=60	$p=\frac{1}{2}$	$p=\frac{1}{2}$
MoB payout	N/A	$1.00*
Total expected payout (per person-per round)	Expected payout =$3.00+$5*0.50 = $5.50 Expected payout-cost of effort (k=0.0002) = $5.50- $2.80 **=$2.70**	Expected payout =$3+$5*0.5+ $1 = $3+$2.50+$1 =$6.50 Expected payout-cost of effort (k=0.0002) = $6.50- $2.80 **=$3.70**

*Man-on-the-bench payouts are imputed to players in the MoB condition, since they are the only ones who can exert effort.

APPENDIX B
Social cohesion measurement scale

Social Cohesion (1 = "strongly disagree," and 7 = "strongly agree")

1. We work well together in our group.
2. I feel that I belong to this group.
3. I am happy to be part of this group.
4. I see myself as part of this group.