Reference Points and Organizational Performance: Evidence from Retail Banking

Douglas H. Frank[*] and Tomasz Obloj[**]

March 10, 2011

Abstract

Economic theories of organizations describe organizational decisions as rational responses to prevailing incentive structures. In contrast, behavioral theories suggest that organizational decisions reflect bounded rationality and cognitive biases. In this paper, we explore and distinguish empirically these two competing views and their performance consequences. We study the daily performance of 164 units of a retail bank throughout a two-month sales tournament. Former tournament leaders—units who have occupied a prize-eligible rank but have fallen out—have 28 percent higher daily sales than units who have never led. This is not due to underlying productivity differences; neither is it fully attributable to the prevailing incentive structure. Rather, outlets appear to be motivated to regain a lost "endowment": the contest ranking entitling them to a prize. Our results therefore suggest that—in addition to the effects predicted by standard economic theories—a behavioral mechanism partially determines the units' performance.

Keywords: Behavioral Economics, Prospect Theory, Reference-Dependent Utility, Tournaments, Aspiration Levels, Behavioral Theory of the Firm, Endowment Effect

[*] Assistant Professor of Strategy, INSEAD, Boulevard de Constance, 77305 Fontainebleau. *Email*: douglas.frank{at}insead.edu.

[**] Ph.D. candidate in Strategy, INSEAD, Boulevard de Constance, 77305 Fontainebleau. *Email*: tomasz.obloj{at}insead.edu.

1 Introduction

Economic theories of organizations are typically theories of rational actors responding optimally to optimally designed incentive systems (Holmstrom & Milgrom 1994). Yet the rational, self-interested "homo economicus" who populates these theories is a controversial figure. Much research documents important ways in which humans appear to be less selfish and more boundedly rational than homo economicus.[1] However, defenders of the classical economic view counter that its critics are too quick to abandon it, and point out that empirical findings that appear to contradict it can actually be explained within the paradigm (Binmore 2005; Plott & Zeiler 2007).

Deviations from the rational actor model have long been a staple of management theory. For example, the behavioral theory of the firm takes bounded rationality as a starting point in analyzing decision making in organizations and organizational behavior (Cyert & March 1963). A central premise in this tradition is that organizations' decisions are influenced by their performance against aspiration levels (see Argote and Greve 2007 for a review). Empirical support for this proposition comes from studies showing that an organization's performance relative to a peer group or to its own historical standards affects choices such as risk taking (Bromiley 1991; Nickel & Rodriguez 2002; Audia & Greve 2006), inertia in strategic actions (D. Miller & Chen 1994), investments (Greve 2003) and tie formation (Baum et al. 2005).

However, we argue that—despite strong empirical evidence of a relationship between organizations' performance against social or historical norms on the one hand and important strategic decisions on the other—a fundamental question is still

[1] Surveys include Rabin (1998), Camerer et al. (2004) and DellaVigna (2009).

left open: are these patterns evidence of a boundedly rational ("behavioral") decision making process, or can they also be explained by a fully rational ("economic") one? For example, one robust finding in the management literature—that performance below aspiration levels increases the probability of risky organizational change—is mirrored by Chevalier and Ellison's (1997) finding that mutual fund managers make riskier investments when the fund's performance lags a market benchmark.[2] Yet Chevalier and Ellison (1997) also show that the managers' behavior is a perfectly rational response to their career concerns. This highlights a crucial point in interpreting the evidence on aspiration levels: without knowledge of the institutional context in which decisions are made—in particular, without knowing the structure of managers' incentives—it is virtually impossible to disentangle the fully rational mechanisms from the boundedly rational ones.

In this paper, we attempt to separate the behavioral and economic mechanisms underlying organizational decision making, and to estimate their relative importance for organizational performance. We start with a theoretical model that builds on Köszegi and Rabin's (2006) formalization of prospect theory (Kahneman & Tversky 1979). The key feature of prospect theory—and our model—is that an individual's well-being is based on her position (e.g., wealth, income, consumption) relative to some reference point, rather than on her absolute position. This is the source of the well-known "endowment effect" made famous by Thaler (1980). Our model predicts that employees performing below their reference point will have superior motivation to work hard than those who feel that they are "ahead". Thus, for a fixed incentive system, organizational performance can be affected by events that shift employees' reference points.

[2] A market benchmark would be characterized as a social aspiration level in the aspiration level literature.

We test the model's predictions using a unique data set from a multiunit retail bank in a large European country. During a two-month period, the bank operated a contest in which employees at the top-performing outlets received all-expenses-paid vacations. Outlets received daily updates on their performance ranking. Because of day-to-day fluctuations in performance, outlets regularly rose and fell in the rankings. Therefore, on any given day, the pool of "followers" (outlets below the minimum rank needed to win the prize) included both (a) former leaders and (b) outlets who had never led. Our model predicts that, if there is an endowment effect (an increase in the reference point) associated with holding the lead ("temporarily winning" the prize), then daily performance should be higher in group (a). This is what we find: "former leadership" implies an increase in daily performance of 28 percent over the average of all outlets.

We find no evidence to suggest that these results are driven by inherent productivity differences across outlets. First, the results are conditional on each outlet's position in the contest, meaning that outlets performing at the same average rate are compared. Second, outlets are "handicapped" by the bank: the performance that matters for the prize is relative to that outlet's past performance, before the contest was announced. When contestants are of equal ability, winners are randomly determined (Lazear & Rosen 1981). Therefore, if handicapping effectively controls for outlet productivity, all outlets should be equally likely ex ante to win the contest. Indeed, we find that pre-contest performance rankings do not predict final rankings in the contest.

Linking our model and the empirical results is the premise that occupying a leading position—crossing the winning threshold—is the event that increases the reference point. To test this, we exploit a "quasi-experiment" in the contest design.

The contest was actually four distinct contests operating in parallel. The minimum rank for a prize varied across contests, from one to four. Thus, there was a one-prize contest, a two-prize contest, and so on; outlets were assigned to only one of these. Comparing the contests with different prize structures, we find a performance increase associated with falling out of the *n*th place in an *n*-prize contest, but no performance increase associated with falling from that same position in a contest with fewer than *n* prizes. This strongly suggests that occupying a prize-eligible rank—temporarily being "endowed" with the prize—is the event that shifts the reference point. Because we compare performance across organizational units operating under a common incentive structure—in contrast to existing work that compares outcomes across diverse organizations with diverse (and unobserved) incentive structures—and because we control for the detailed structure of managers' incentives, the results point clearly to the type of decision making bias proposed in behavioral theories of the firm.

Wherever deviations from standard economic models of behavior affect organizational outcomes, organizational performance may be improved if the design of structures and incentives takes these deviations into account. For example, tournament theory predicts the optimal number of winners in contests such as the one we study (Gibbs 1996). However, our results show that there is an incentive effect associated with temporarily winning the prize that cannot arise in a standard tournament model. This implies, for instance, that the optimal number of prizes in a contest may be higher than tournament theory predicts, if having more prizes increases the probability that non-winners are at least temporary winners at some point.

The rest of the paper is organized as follows. Section 2 presents the model and develops our main propositions. Section 3 discusses the institutional setting, the structure of the sales contest and the data. Section 4 describes the estimation procedure, and Section 5 discusses the results. Section 6 concludes.

2 Model

We assume an organization—which could be a firm or a unit within a firm—whose performance depends on a manager's effort. The manager's pay is increasing in performance. The following describes the model's primitives:

effort:	e;
output:	$y = e + \varepsilon$; ε has distribution F;
wage:	$w = w(y)$;
manager utility:	$U = u(w) + \mu(w - r) - c(e)$.

Assume that $w'(y) > 0$ and $w''(y) \leq 0$. The utility function U consists of a strictly concave utility function, $u(w)$; a strictly convex cost of effort function, $c(e)$; and a gain-loss utility function, $\mu(w - r)$, in which r is the reference point (Köszegi & Rabin, 2006). We consider two alternative sets of assumptions for μ:

i. Standard Preferences:
   a. $\mu(x) = 0$;
   b. $\mu'(x) = 0$;

ii. Nonstandard Preferences:
   a. $\mu'(x) > 0$;
   b. if $x > 0$, then $\mu'(-x) - \mu'(x) > 0$;
   c. $\mu''(x) = 0, \forall x \neq 0$.

Assumption ii.b. is loss aversion. Assumption ii.c. rules out diminishing sensitivity in order to focus on the consequences of reference dependence and loss aversion. It also eliminates the possibility of preference reversals for stochastic consumption bundles (Köszegi & Rabin 2006). Assume that the manager selects effort to maximize expected utility, $E[U]$.

Lemma: Under standard preferences, expected output, $E[y]$, is independent of r.

Proof: From assumption i.b., the manager's utility and, therefore, effort choice is independent of r.■

Proposition: Under nonstandard preferences, $r' > r$ implies that $E[y \mid r'] > E[y \mid r]$. That is, expected output is increasing in the reference point.

Proof: For arbitrary *r*, define $e^*(r)$ as

$$\arg\max_e \int [u(w(e+\varepsilon)) + \mu(w(e+\varepsilon) - r)]dF - c(e) \qquad (1.1)$$

By the functional form assumptions, the objective function is concave and $e^*(r)$ is characterized by:

$$\int [u'(w(e+\varepsilon)) + \mu'(w(e+\varepsilon) - r)]w'(e+\varepsilon)dF = c'(e). \qquad (1.2)$$

Define $\varepsilon_0(r) \equiv \varepsilon \ s.t. \ \mu(w(e+\varepsilon) - r) = 0$. In other words, ε_0 is the value of the random component that equates the realized wage and the reference point, for a given effort choice. That $r' > r$ implies that $\varepsilon_0(r') > \varepsilon_0(r)$. Assumption ii.b. then implies that, for ε in the interval $(\varepsilon_0(r), \varepsilon_0(r'))$, $\Delta \equiv \mu'(w(e+\varepsilon) - r') - \mu'(w(e+\varepsilon) - r) > 0$. Assumption ii.c. implies that, for ε outside this interval, $\mu'(w(e+\varepsilon) - r') = \mu'(w(e+\varepsilon) - r)$. The condition characterizing $e^*(r')$ is therefore:

$$\int [u'(w(e+\varepsilon)) + \mu'(w(e+\varepsilon) - r)]w'(e+\varepsilon)dF + \int_{\varepsilon_0(r)}^{\varepsilon_0(r')} \Delta w'(e+\varepsilon)dF = c'(e).$$

(1.3)

From inspection of (1.2) and (1.3), $e^*(r') > e^*(r)$. The result follows. ■

Figure 1 illustrates the proof. Wage realizations are on the horizontal axis. Conditional on effort, the wage is a random variable and so the horizontal axis depicts the support of a distribution that is induced by *F*. This distribution is not illustrated because the results hold for any arbitrary distribution. The effort choice equates the marginal cost of effort with the expected marginal utility. The expected marginal gain-loss utility is the average slope of each curve. Except for wage realizations in the shaded region—the interval corresponding to $(\varepsilon_0(r), \varepsilon_0(r'))$—this is the same for r and r'. In the shaded region, the slope of the dashed curve—corresponding to the higher reference point—is greater than the slope of the solid curve. Therefore, for any distribution *F*, the expected marginal gain-loss utility increases with the reference point. Because only the gain-loss utility is affected by changes in the reference point, the optimal effort must increase and expected output with it.

Insert Figure 1 around here

Before proceeding to the empirical analysis, we note that the reference point in this model is an exogenous parameter; we do not propose a theory of how the reference point is formed. Rather, the model is presented to sharpen a very simple intuition: if individuals' preferences display reference dependence and loss aversion, they will work harder when they are behind their reference point than when they are ahead, and this will be observable in organization-level performance.

3 Empirical Context and Data

We test our main proposition using a confidential dataset that contains detailed information on all outlets of a private retail bank in a large European country. The bank is among the twenty largest financial institutions in the country, employing several thousand people and serving hundreds of thousands of customers. Its focus is on sales of simple financial products, such as deposit accounts and small personal loans, to mass market customers. The bank operates through a network of standardized outlets located in large to mid-size towns. A typical outlet employs three to four salespeople. The institution we study has therefore a typical multi-unit structure.

This dataset is well-suited to test our theory for three main reasons. First, it contains daily, outlet-level results, enabling us to observe daily changes in performance across outlets as the tournament unfolds. Second, it contains information on *all* outlets of the bank over the *entire* period of the tournament. Hence, it captures all longitudinal and cross-sectional variation without suffering from sample selection bias, attrition, or censoring. Third, because it is characterized by high-powered incentives (see, for example, Hubbard & Palia 1995 and Chevalier & Ellison 1997), the financial services industry is an ideal setting to study the factors that influence responses to incentives.

3.1. Sales Contest

There were 164 outlets participating in a two-month sales contest. Outlets were ranked according to the number of primary personal loans[3] sold over the contest period. All employees at the top-ranked outlets received a one week holiday at an exotic resort, paid by the bank. Outlets were "handicapped" in the sense that each outlet's ranking was based on its quantity of loans sold divided by its own monthly average in the four months preceding the contest. Because the bank announced the

[3] While the bank sells many types of financial products, personal loans are the most important, accounting for over 90% of pre-tax profits during our observation period. According to the Sales Director, "What we sell are personal loans. Personal loans are where we make money. If we sell anything else, it is so that we can sell more personal loans." Personal loans are of two types: *primary*—loans sold to first-time customers with the bank— and *secondary*—loans sold to returning customers (typically, with a positive history of repayment of the primary loan). Primary loans account for over 50% of total sales and over 70% of pre-tax profits from personal loans during the period we study. The contest was over primary loans only.

contest just 4 days before it began, outlets had little opportunity to influence their performance benchmark.

Bank outlets did not have a formal choice with respect to participating in the tournament. While this feature of our data does not allow us to study issues of self-selection (Lazear & Rosen 1981), it is invaluable for our line of inquiry as there is no attrition among tournament participants which could affect the results. The closest study to ours with respect to the empirical setting—Casas-Arce and Martinez-Jerez (2009)—suffered from such attrition of participants which prevented the authors from fully investigating of the dynamic properties of tournament participation.

In the contest, each outlet was assigned to one of four groups of equal size. Assignment was on the basis of the performance benchmark—the number of primary loans that outlet sold in the four months preceding the tournament. The number of prizes available (i.e., the number of outlets who could win the holiday) varied by group: Group 1 (outlets with the lowest benchmark) competed for one prize; Group 2 for two prizes, and so on up to four.[4] Each day, 1 hour prior to the official earliest opening time, the interim contest results (rank and performance of all outlets) were distributed to outlet managers. Therefore, all outlets had full information about their own and competitors' performance throughout the duration of the tournament. Importantly, because the number of prizes was centrally assigned by the headquarters, it is not a choice variable for the outlets that we study. This is an important feature of our data because prior research has argued that the strength of incentives may affect self-selection of employees along their risk attitudes (i.e., organizations with stronger

[4] The bank's rationale was that, because the outlets with higher benchmarks generated higher average profits, it was appropriate to offer more prizes. This is broadly consistent with theoretical prescriptions that the optimal incentive strength is increasing in the marginal productivity of effort (Holmstrom & Milgrom 1991). Gibbs (1996) predicts that incentive strength is increasing in the number of tournament prizes (up to a limit which is not reached in the present context).

incentives may naturally attract managers who are less risk averse) (Dushnitsky & Shapira 2010). Our sample does not suffer from these possible selection effects.

The main data set consists of 9053 outlet-day observations. This is lower than the theoretical maximum of 10,004 (61 calendar days and 164 outlets) because of weekends and holidays during which some of the outlets were closed. On average, outlets were open for business 49 days during the contest, with a maximum of 58 and a minimum of 39. For some of the robustness checks described below, we use available data from before and after the tournament.

3.2. Variables

We study two measures of performance outcomes. The first, *output,* is the number of primary personal loans sold by an outlet on a given day. Because the tournament performance measure is sales divided by an outlet-specific benchmark, in some specifications we measure daily output in an analogous fashion. The results are qualitatively unchanged. As previously discussed, another moment of the distribution of performance—its variance—has been associated with theories of reference-dependent utility. Although this is not the focus of our inquiry, we explore this for comparison with existing work. We measure the *variance* as the squared deviation of daily outlet performance from the outlet mean. We define the variance with respect to both (a) mean output over the duration of the tournament; and (b) mean-output over a 5-day trailing window, as is common in the finance literature (Kolasinski 2009).

Our main independent variable concerns the shift of the reference point of the tournament participants. We hypothesize that, if an outlet occupies a prize-eligible rank at any point during the contest, winning a prize becomes its reference point. Because of day-to-day fluctuations in outlet performance—caused both by random demand variation and exogenous differences in the days on which outlets are open—

occupants of the prize-eligible ranks constantly change. We define a *former leader* as an outlet which, at date *t*, does not occupy a prize-eligible rank but has done so in the past.[5] Our model predicts that former leaders will have higher daily performance than other outlets at a similar position in the tournament..

Our additional independent variables are linked to the dynamic nature of the tournament. As shown by Casas-Arce and Martínez-Jerez (2009), tournament participants will alter their effort based on their distance from the current leaders, as this affects the perceived probability of winning the tournament and, therefore, the expected returns to effort. Effort as a function of relative tournament position has an inverted U-shape: contestants at the top of the rankings "coast"—since winning seems assured—participants at the bottom give up, and those in between work the hardest. To account for these effects we introduce two variables: *leading distance* and *trailing distance.* The former is defined as $leading\ distance_{i,t} = \max\{0, P_{i,t-1} - P_{j,t-1}\}$, where *P* is cumulative output (relative to the benchmark—i.e., the measure on which the rankings are based) and *j* indexes the outlet occupying the lowest prize-eligible rank. Therefore, this distance is zero for the lowest-ranked current leader and all followers. The trailing distance is defined analogously[6].

The sales contest we study was a temporary incentive in addition to an ongoing incentive program in which outlet employees received bonuses for sales of personal loans. The bonus rate varied with the level of performance against a monthly target, meaning that the immediate marginal returns to effort varied with progress against the target. To control for possible confounding effects of this second incentive, we therefore include the variable *bonus progress,* measured as the outlet's exact position

[5] Note that this indicator variable takes a zero value for any days in which a former leader regains a prize-eligible rank.
[6] It is important to note that while our distance variables may reflect the perceived probability of winning the tournament, in this study we are not explicitly interested in exploring the role of subjective estimates of probabilities on outcomes (March & Shapira 1987).

with respect to its sales target on a given day. Our results are robust to an alternative specification of this control—the outlet's actual bonus rate for the marginal loan sold.

We exploit the fact that outlets were assigned to different contests with different prize structures to introduce another control variable, based on simulating the structure of the four-prize tournament on all other outlets. We define the variable *former Top-4,* which equals one if an outlet would be a former leader under the rules of the four-prize tournament. Including this variable alongside the former leader variable is a regression discontinuity test of sorts that permits us to compare the performance of similarly-placed outlets across tournaments operating under different rules. If the reference point is affected simply by being "near the top", then the former Top-4 variable will predict performance. In contrast, if it is crossing the prize threshold at any point that shifts the reference point, then only the former leader variable will predict performance.

In our models we also include outlet level fixed effects (see estimation details below) to allow for outlet-specific unobservable characteristics that could drive the responses to organizational incentives as well as demand for loans. Table 1 details the summary statistics and correlations for all variables.

Insert Table 1 around here

----

4 Estimation

Our basic empirical model takes the following form:

$$Y_{i,t} = \alpha_0 + \alpha_1 Y_{i,t-1} + \alpha_2 X_{i,t} + \varphi R_{i,t-1} + \beta Z_{i,t} + u_i + e_{i,t} \tag{1.4}$$

where *Y* corresponds to the performance variable (*Output* or *Variance*), *X* is a vector of independent variables affecting performance and *R* is an indicator variable indicating a former leader (an outlet whose performance trails its reference point). *Z* is a vector of additional control variables. A major concern with the data we study relates to the possible autocorrelation in shocks to performance, a well established property of time-series sales data (Kapoor, Madhok, & Wu, 1981). Because the controls in *X* include cumulative output, the model implicitly includes lagged dependent variables, and therefore we include the lagged value $Y_{i,t-1}$ explicitly in (1.4) . A model of this form, including the lagged dependent variable and individual fixed effects, is biased by construction when estimated with ordinary least squares regression (Nickell 1981). In order to obtain unbiased estimates, we use the "difference GMM" estimator in the form proposed by Arellano and Bond (1991). This estimation strategy is similar to the one used by Casas-Arce & Martínez-Jerez (2009) to analyze a tournament with a structure resembling ours..

The identification of this model relies on first-differencing the model (in order to remove the fixed effects) and using lagged values of independent variables as their instruments. Given that all our independent and control variables are a function of outlet's past performance, we treat them as predetermined in our models. In all our specifications, we find significant negative first-order serial correlation, while there is no evidence of second-order autocorrelation in first-differenced residuals. This indicates that the disturbances in the differenced specification are not serially correlated, which is essential for the consistency of the estimation procedure. Similarly, the Sargan statistic is well below its critical levels in all models, implying joint validity of the moment conditions[7].

[7] The lowest *p* value we observe is 0.52 in Model 1.

5 Results

The descriptive results in Table 2 preview the main results of the analysis. Former leaders—outlets who do not currently occupy a prize-eligible rank but have done so in the past—have higher average daily output than other outlets. Note that these other outlets include the former leaders on the days (a) when they occupy a prize-eligible rank; and (b) before they first occupied a prize-eligible rank. Former leaders also have higher daily output variance, whether measured with respect to their long-term average or a five-day moving window.

Insert Table 2 around here

5.1 Output

Table 3 presents the difference GMM estimates with controls, to take into account the factors that might lead the simple descriptive statistics to be misleading. Columns 1 and 2 confirm the basic predictions of Casas-Arce and Martinez-Jerez (2009): output as a function of position in the tournament has an inverted-U shape. In fact, we find generally more robust support for this hypothesis than in Casas-Arce and Martinez-Jerez (2009). This may confirm those authors' conjecture that their results are affected by sample selection and attrition biases, which cannot arise in our setting.

Insert Table 3 around here

In column 3 we add the control for bonus progress—where the outlet stands with respect to its monthly sales target. Increasing values of this variable signify a higher bonus rate for the marginal loan sold under the standard incentive plan that was in effect alongside the tournament. This control is positive and significant, and remains so under all subsequent specifications.

In column 5, we introduce the former Top-4 control, indicating that the outlet held a Top-4 position in the past but not on the date observed. Column 4 shows that former Top-4 outlets have greater output. However, when the former leader variable is added in column 5, the former Top-4 variable loses significance and changes sign. Observe that, under the contest rules for Group 4, the former Top-4 variable is synonymous with the former leader variable. However, for Groups 1-3, the two need not coincide. Conditional on former Top-4, variations in former leader separate the outlets who actually crossed the winning threshold in the past from those who narrowly missed it. Column 5 is therefore a regression discontinuity analysis of sorts. If the reference point is affected simply by being "near the top", then the coefficient on former Top-4 should be positive and significant. That it is neither suggests a qualitative change that comes with temporarily being "endowed" with the prize.

Columns 6 and 7 introduce time controls (contest period divided into four equal stages), alone and interacted with the distance measures. The specification in column 7 therefore allows for outlets' output to vary depending on their time horizon (time remaining in the contest) and allows their response to their position in the contest to vary with this horizon. In other words, the results in column 7 tell us that, controlling for cumulative performance (the distance measures) and time remaining in the contest, outlets whose remaining difference is that they were once tournament leaders sell more loans on average—as our theory predicts.

One possible concern with our results is that we use absolute output levels as our dependent variable, even though performance in the tournament is measured relative to an outlet-specific benchmark. Therefore, in column 8 we present results based on the latter measure. Neither the statistical significance nor the direction of our results is affected.

Apart from the statistical significance of the results, Table 3also reveals important economic effects. The following analysis is based on results presented in column 8. Regarding the dynamic tournament incentive effects, an outlet at the mean level of trailing distance has daily output of 61 percent below the sample mean, indicating that being even moderately behind the prize threshold has a strong disincentive effect. An outlet at the mean level of leading distance has daily output 3.5 below the sample mean, indicating that being ahead of the prize threshold also has a demotivating effect. (The effect is smaller than for the trailing distance because there are at most four leaders in any of the tournaments and so the range of leading distances is much smaller.)

The effect of a change in the reference point is also sizable. Outlets who have fallen from leading positions have daily output 28 percent above the sample mean. To put this figure in perspective, imagine two outlets at the mean distance behind the prize threshold. The demotivating effect of trailing the leaders (61 percent daily output reduction) is roughly halved if the outlet has itself been a leader in the past.

5.2 Output Variance

As noted above, previous research has associated risk-seeking behavior with organizations' falling short of their "aspiration levels", and used these results to infer a mechanism based on reference-dependence (Audia & Greve 2006; Greve 1998). Also, a large body of research on behavioral decision theory has investigated individual managers' deviations from standard expected utility theory as manifested in risk attitudes (March & Shapira 1987; Shapira & Venezia 1992; K. D. Miller & Shapira 2004). Although such behavior is not a focus of our theoretical analysis, our empirical setting is ideal to study the link between non-standard preferences and risk-taking. Results of this analysis are presented in Table 4.

Insert Table 4 around here

With respect to the former leader variable, the same broad patterns emerge as in Table 3: whether measured with respect to the long-term average or a five-day moving average, daily output variance increases for outlets who are former leaders. This result is robust to multiple controls. In contrast to Table 3, none of the distance measures is a significant predictor of sales variance, although the signs mirror those for the output analysis. The broad conclusion is that high-variance sales strategies are motivated more by an attempt to regain a lost, winning position than by an attempt to keep pace with the leaders.

5.3 Alternative Hypotheses

Because an outlet cannot become a former leader until it has led at some point, one might imagine that this could produce a mechanical correlation between former leader status and daily output. This is unlikely, for several reasons. First, the results in Table 3 include controls for position in the contest (leading and trailing distance) and their interaction with time. Therefore, the regressions compare outlets which have the same cumulative performance and the same time horizon, and which differ only by the fact of being former leaders or not. Second, column 5 in Table 3 shows that there is an effect of being a former leader that is distinct from merely having been near the top of the rankings. The results show that all of the explanatory power in the former Top-4 variable is coming from the outlets that were actually former leaders. There is no significant impact on output of having come close, and in fact the point estimate is the opposite sign.

Finally, if the former leader variable has a mere mechanical relationship to performance, we should find this relationship during periods when the contest was not in effect. We therefore performed two additional analyses, simulating the contest on each of the 61-day periods before and after the actual contest period—applying the exact rules and performance measures in effect during the true contest. In both cases, the former leader variable does not predict output.

Another possible explanation for our results is that outlets who are former leaders just have higher productivity on average. This is unlikely, for several reasons. First, as already discussed, the controls for distance to the leaders and time mean that outlets with similar cumulative performance at similar stages of the contest are compared in the regressions. Second, outlets' performance in the contest is measured with respect to their average output in the four months preceding the contest. By design, the contest "handicaps" outlets to control for inherent productivity differences (due, for example to differences in local demand characteristics) and thereby give them an approximately equal chance at winning. This is consistent with the predictions of tournament theory that, among contestants of equal ability, winners are randomly determined (Lazear and Rosen, 1981). Table 5 supports this hypothesis: there is no significant correlation between outlets' ranking according to their performance benchmarks and their final ranking in the tournament.

Insert Table 5 around here

A further alternative hypothesis relates to our choice of the output measure. While the structure of the incentives indicates that the number of personal loans is a good measure of outlet performance, the true objective function of the bank is unknown to us. Other possibilities are that the organization is trying to maximize either the volume of sales of loans or profits on loans. The direction and significance of our results remains under either alternative specification.

The final robustness check pertains to our construction of the distance measures. We use differences in cumulative output according to the contest measure (i.e., relative to the past performance benchmark) to measure the leading and trailing distance controls, assuming that it is this distance that affects outlets' perceived probability of winning the contest and, therefore, their output. An alternative hypothesis is that outlets respond not to the cardinal distance but rather the ordinal ranking. Our results are qualitatively unchanged under this alternative specification.

6 Conclusion

In this paper, we develop and test a model of organizational performance when managers have reference-dependent preferences. Our model predicts that, holding explicit incentive effects constant, increases in the reference point will be associated with increases in performance. We test this prediction using detailed data from a sales contest at a multiunit retail bank. We show that daily output is greater among outlets which have held a prize-eligible contest rank in the past, even when controlling for other important determinants of sales effort. Moreover, outlets who have come close to, but not attained, the prize-eligible ranks and fallen back show no similar increase in performance, suggesting that there is an endowment effect associated with holding a prize-eligible rank, even temporarily. Our results are not consistent with alternative hypotheses—such as underlying productivity differences across outlets or mechanical correlations between average output and attaining a high rank—because the contest design controlled for productivity differences, and simulating the tournament rules on non-tournament periods fails to produce the same patterns. Our results provide rare evidence that cognitive biases that are well-established at the individual level have an impact on organizational outcomes.

References

Arellano, M. & Bond, S., 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277-297.

Argote, L. & Greve, H.R., 2007. A Behavioral Theory of the Firm--40 Years and Counting: Introduction and Impact. *ORGANIZATION SCIENCE*, 18(3), 337-349.

Audia, P.G. & Greve, H.R., 2006. Less Likely to Fail: Low Performance, Firm Size, and Factory Expansion in the Shipbuilding Industry. *MANAGEMENT SCIENCE*, 52(1), 83-94.

Baum, J.A.C. et al., 2005. Dancing with Strangers: Aspiration Performance and the Search for Underwriting Syndicate Partners. *Administrative Science Quarterly*, 50(4), 536-575.

Binmore, K., 2005. Economic Man – or Straw Man? *Behavioral and Brain Sciences*, 28(06), 817-818.

Bromiley, P., 1991. Testing a Causal Model of Corporate Risk Taking and Performance. *The Academy of Management Journal*, 34(1), 37-59.

Camerer, C., Loewenstein, G. & Rabin, M., 2004. *Advances in behavioral economics*, Princeton University Press.

Casas-Arce, P. & Martinez-Jerez, F.A., 2009. Relative Performance Compensation, Contests, and Dynamic Incentives. *MANAGEMENT SCIENCE*, 55(8), 1306-1320.

Chevalier, J. & Ellison, G., 1997. Risk Taking by Mutual Funds as a Response to Incentives. *The Journal of Political Economy*, 105(6), 1167-1200.

Cyert, R.M. & March, J.G., 1963. *Behavioral Theory of the Firm*, Wiley-Blackwell.

DellaVigna, S., 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47(2), 315-372.

Dushnitsky, G. & Shapira, Z., 2010. Entrepreneurial finance meets organizational reality: comparing investment practices and performance of corporate and independent venture capitalists. *Strategic Management Journal*, 31(9), 990-1017.

Gibbs, M., 1996. Promotions and Incentives. *Working Paper, University of Chicago Graduate School of Business*.

Greve, H.R., 1998. Performance, Aspirations, and Risky Organizational Change. *Administrative Science Quarterly*, 43(1), 58-86.

Greve, H.R., 2003. Why are there So Many Multiunit Organizations? *Strategic Organization*, 1(1), 109 -115.

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

Holmstrom, B. & Milgrom, P., 1994. The Firm as an Incentive System. *The American Economic Review*, 84(4), 972-991.

Hubbard, R.G. & Palia, D., 1995. Executive pay and performance Evidence from the U.S. banking industry. *Journal of Financial Economics*, 39(1), 105-130.

Kahneman, D. & Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291.

Kolasinski, A.C., 2009. Subsidiary debt, capital structure and internal capital markets. *Journal of Financial Economics*, 94(2), 327-343.

Köszegi, B. & Rabin, M., 2006. A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics*, 121(4), 1133-1165.

Lazear, E.P. & Rosen, S., 1981. Rank-Order Tournaments as Optimum Labor Contracts. *The Journal of Political Economy*, 89(5), 841-864.

March, J.G. & Shapira, Z., 1987. Managerial Perspectives on Risk and Risk Taking. *Management Science*, 33(11), 1404-1418.

Miller, D. & Chen, M., 1994. Sources and Consequences of Competitive Inertia: A Study of the U.S. Airline Industry. *Administrative Science Quarterly*, 39(1), 1-23.

Miller, K.D. & Shapira, Z., 2004. An empirical test of heuristics and biases affecting real option valuation. *Strategic Management Journal*, 25(3), 269-284.

Nickel, M.N. & Rodriguez, M.C., 2002. A review of research on the negative accounting relationship between risk and return: Bowman's paradox. *Omega*, 30(1), 1-18.

Nickell, S., 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6), 1417-1426.

Plott, C.R. & Zeiler, K., 2007. Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory? *The American Economic Review*, 97(4), 1449-1466.

Rabin, M., 1998. Psychology and Economics. *Journal of Economic Literature*, 36(1), 11-46.

Shapira, Z. & Venezia, I., 1992. Size and frequency of prizes as determinants of the demand for lotteries. *Organizational Behavior and Human Decision Processes*, 52(2), 307-318.

Thaler, R., 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39-60.

Table 1

Summary Statistics

	Variable	Mean	S.D	1	2	3	4	5	6	7	8	9
1.	Output	4.23	1.99									
2.	Output variance (long-term avg)	3.20	6.47	0.51								
3.	Output variance (running avg)	2.64	5.07	0.48	0.85							
4.	Trailing distance	6.00	5.72	-0.13	-0.10	-0.11						
5.	Trailing distance squared	68.76	122.74	-0.09	-0.08	-0.08	0.87					
6.	Leading distance	0.10	0.51	0.02	0.03	0.04	-0.21	-0.11				
7.	Leading distance squared	0.27	2.73	0.00	0.02	0.03	-0.10	-0.06	0.82			
8.	Bonus progress	0.44	0.31	0.15	0.07	0.07	-0.04	-0.03	0.02	0.01		
9.	Former Top-4	0.32	0.43	0.00	0.04	0.05	-0.45	-0.33	0.29	0.14	0.06	
10.	Former leader	0.21	0.38	0.06	0.01	0.02	-0.43	-0.28	0.35	0.17	0.05	0.74

Table 2

Reference Point Effects: Descriptive Statistics

	Former Leaders	**Others**	**Difference**
Output	4.31	4.20	0.11**
Output variance (long-term average)	3.34	3.13	0.21**
Output variance (running average)	2.74	2.58	0.16**

Notes: **Significant at 0.05.

Table 3

Reference Point Effects on Output

	DV: Daily Performance—Absolute							DV: Perf./Benchmark
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Lagged output	0.02	0.02	0.005	0.010	0.002	0.004	0.002	0.013
	(0.014)	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)
Trailing distance	0.515***	0.612***	0.576***	0.528***	0.716***	0.583***	0.554***	0.011***
	(0.058)	(0.071)	(0.070)	(0.067)	(0.072)	(0.067)	(0.081)	(0.001)
Trailing distance squared		-0.019***	-0.014**	-0.013**	-0.014***	-0.009*	-0.009	0.0001*
		(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.00)
Leading distance	-0.313***	-0.873***	-1.048***	-1.014***	-0.926**	-1.033***	-1.249***	-0.029***
	(0.099)	(0.024)	(0.282)	(0.202)	(0.204)	(0.202)	(0.276)	(0.006)
Leading distance squared		0.075***	0.089***	0.085***	0.082***	0.085***	0.104***	0.002***
		(0.023)	(0.023)	(0.023)	(0.024)	(0.023)	(0.028)	(0.00)
Bonus progress			0.22**	0.388***	0.449***	0.791*	0.530*	0.028*
			(0.102)	(0.126)	(0.124)	(0.398)	(0.299)	(0.015)
Former Top-4				1.279**	-1.385	-0.496	-0.057	0.002
				(0.560)	(0.918)	(0.866)	(0.641)	(0.02)
Former leader					2.792***	2.421***	2.112***	0.03***
					(0.635)	(0.793)	(0.788)	(0.005)
Time controls						Included	Included	Included
Time controls * trailing							Included	Included

distance								
Time controls * leading distance							Included	Included
AR(1)	-60.69***	-61.01***	-60.06***	-60.68***	-56.46***	-59.15***	-58.96***	-59.11***
AR(2)	1.37	1.47	0.51	0.69	0.38	1.01	0.89	1.45
Sargan	160.93	186.29	185.28	186.79	184.89	179.85	181.21	178.64

Notes: *p<0.1, **p<0.05, ***p<0.01; N=8667 for all models. Constant included but not reported.

Table 4

Reference Point Effects on Output Variance

	DV: Variance compared to mean					DV: Variance over running window				
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**	**Model 10**
Lagged variance	-0.003	-0.002	-0.003	-0.003	-0.004	-0.026*	-0.026*	-0.028*	-0.029*	-0.033**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)
Trailing distance	0.275	0.241	0.186	0.194	0.257	0.257	0.140	0.111	0.114	0.208
	(0.24)	(0.23)	(0.24)	(0.23)	(0.24)	(0.198)	(0.18)	(0.19)	(0.20)	(0.189)
Trailing distance squared	-0.014	-0.021	-0.019	-0.022	-0.004	-0.008	-0.013	-0.008	0.002	0.005
	(0.02)	(0.024)	(0.023)	(0.021)	(0.02)	(0.019)	(0.02)	(0.019)	(0.02)	(0.017)
Leading distance	-0.678	-0.699	-0.740	-0.732	-0.348	-0.630	-0.642	-0.683	-0.666	-0.064
	(0.71)	(0.71)	(0.712)	(0.711)	(0.972)	(0.71)	(0.71)	(0.712)	(0.713)	(0.88)
Leading distance	0.11	0.105	0.111	0.112	0.080	0.09	0.09	0.087	0.083	0.123
Squared	(0.08)	(0.08)	(0.08)	(0.08)	(0.10)	(0.15)	(0.15)	(0.16)	(0.15)	(0.16)

Bonus progress	1.05**	1.44***	1.30***	1.89*	1.52	0.328*	0.764**	0.651**	0.221	0.132
	(0.45)	(0.43)	(0.44)	(0.98)	(0.90)	(0.18)	(0.34)	(0.31)	(0.15)	(0.09)
Former Top-4		2.37**	2.00	1.22	0.53		3.45***	2.34*	2.14	1.47
		(1.12)	(2.75)	(1.38)	(0.63)		(1.23)	(1.35)	(1.32)	(1.31)
Former leader			3.44**	2.65**	3.33**			2.63**	2.28**	2.37*
			(1.53)	(1.21)	(1.55)			(1.29)	(1.10)	1.38
Time controls				Included	Included				Included	Included
Time controls * trailing distance					Included					Included
Time controls * leading distance					Included					Included
AR(1)	-63.32***	-63.22***	-63.07***	-63.38	-63.19***	-58.64***	-54.24***	-58.07***	-58.18***	-58.00
AR(2)	-0.45	-0.46	-0.50	-0.42	0.50	-1.39	-1.32	-1.31	-1.21	-1.44
Sargan statistic	180.63	180.57	177.59	170.12	166.55	185.29	183.31	185.29	178.86	175.18

In models 1-3, N=8667. In models 4-6, N=7911. Constant included but not reported.

Table 5

Correlation of Contest and Past Performance

	DV: Final Contest Rank
Pre-contest rank	0.085
	(0.072)
Constant	20.08***
	(2.64)
Contest group dummies	Included
F-statistic	0.40

Notes: N=164. Robust standard errors in parenthesis. Rank order correlation test is insignificant (p>0.40).

Figure 1

Illustration of Proof of Proposition