



Reluctant to change: Self-enhancing responses to diverging performance measures ☆

Pino G. Audia *, Sebastien Brion

Haas School of Business, University of California, Berkeley, CA 94720-1900, USA

Received 8 April 2005

Abstract

Although there is extensive evidence that past performance influences the propensity to make changes, research on how decision makers respond to diverging performance measures has been sparse. This paper addresses this gap in an experimental and a field study in which we examine how decision makers respond to the ambiguity introduced by two diverging performance indicators of unequal importance. Both studies suggest that decision makers respond to diverging performance indicators in a self-enhancing manner. Decision makers gave importance to a secondary performance indicator only when it helped them maintain a sense of positive performance, that is, when a secondary performance measure was high and a primary performance measure was low. The results suggest that, in contexts in which decision makers are likely to experience diverging performance indicators, perceptions of success and the associated reluctance to make changes might be more pervasive than is often thought.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Decision making; Self-enhancement; Change; Performance feedback; Ambiguity

USA Today Interviewer, Ron Insana: Mr. Eisner, the last time Disney faced a takeover threat, corporate raiders, unhappy board members, and a flagging stock price was 1984, and you were brought in to fix things. Why shouldn't that type of change happen now?

Eisner: In 1984, the studio did not come off the biggest year in the history of the motion pictures business. In 1984, the company did not have the leading sports network, ESPN. In 1984, the company did not have theme parks all around the world. In 1984, the company did not have international operations with the Disney Channel. I could go on. The position of the company has never been stronger. The balance sheet has never been stronger. Cash flow has never been stronger. The only

relationship between 2004 and 1984 is that they both have fours in them. (Insana, 2004)

Introduction

This excerpt from an interview with Disney's Chief Executive Officer (CEO) reflects two insights that deserve the critical attention of researchers interested in the relationship between performance and change. First, the excerpt recognizes that the availability of multiple and diverging performance measures contributes to the difficulty in determining unequivocally whether an organization is performing well or poorly. While the interviewer refers to performance indicators that suggest that Disney is experiencing a performance crisis, Eisner calls attention to outcomes that suggest that the position of the company has never been stronger. Second, the excerpt recognizes that the way in which people interpret the ambiguity created by diverging performance indicators is likely to affect their propensity to make

☆ We thank Chip Heath, Barry Staw, and Phil Tetlock for their helpful suggestions.

* Corresponding author. Fax: +1 510 642 4700.

E-mail address: audia@haas.berkeley.edu (P.G. Audia).

changes. While the interviewer asks why the slump in performance should not lead to the same radical changes experienced by the company in the past, Eisner does not see a crisis and probably does not see the need to replace himself.

Understanding how the ambiguity introduced by the availability of diverging performance indicators influences the relationship between performance and change is the focus of this paper. Specifically, we address the following questions: Are decision makers biased in their evaluation of diverging performance measures? Do they give greater attention to indicators signaling success or to indicators signaling failure? Does a biased evaluation of diverging performance measures affect their propensity to make changes?

The availability of multiple and weakly linked performance indicators is a pervasive feature of organizations. Organizational members routinely assess and are assessed using multiple indicators. Furthermore, because the function of multiple performance indicators is to capture differing aspects of performance, these indicators are often contradictory (Meyer, 2002; Meyer & Gupta, 1994). Business Week, for example, ranks companies in the Standard & Poor's (S&P) 500 by using eight criteria of success that often diverge considerably (Business Week, 2004). Similarly, Research and Development managers are aware that various indicators provide differing information regarding not only the performance of their unit but also the performance of individual engineers and scientists (Hauser, 1998).

Surprisingly, the existence of multiple performance indicators is often overlooked in studies of the effect of performance on change. There is evidence both in experimental (e.g., Audia, Locke, & Smith, 2000; Lant & Montgomery, 1987; Marinova, 2004) and field settings (e.g., Boeker, 1997; Greve, 1998; Lant, Milliken, & Batra, 1992) that performance above a reference point (hereafter also called high performance) decreases the probability of change, whereas performance below a reference point (hereafter also called low performance) increases it. Most of this evidence, however, obscures the potential effect of diverging performance indicators because researchers select a priori the performance measure thought to be critical to the individuals or the organizations under investigation.

Despite the lack of empirical studies, two theoretical perspectives suggest conflicting predictions on how decision makers respond to the ambiguity introduced by diverging performance indicators. The first perspective, associated with the behavioral theory of the firm (Cyert & March, 1963; Greve, 2003a) and control theory (Carver & Scheier, 1981; Klein, 1989; Lord & Levy, 1994; Miller, Galanter, & Pribram, 1960; Vancouver, 2005), suggests that organizational members resolve the ambiguity arising from conflicting performance measures by giving greater importance to

those that fall below the reference point. According to this view, low performance measures receive greater attention because individuals are generally motivated by the desire to reduce negative discrepancies between current outcomes and desired outcomes. The second view, associated with research on self-enhancement (Johns, 1999; Sedikides & Strube, 1997; Taylor & Brown, 1988), predicts that individuals will give greater importance to performance measures that are above the reference point. This occurs because individuals tend to be motivated by the desire to protect their self-image from negative evaluations. This literature suggests that, under conditions of ambiguity, such as when information can be interpreted in different ways, individuals display a tendency to engage in self-assessments that are self-serving (Dunning, Meyerowitz, & Holzberg, 1989; Farh & Dobbins, 1989; Huber, 1991).

In this paper, we draw on these two views to investigate how decision makers respond to diverging performance indicators in an experimental and a field study. Because both perspectives assume that decision makers are biased in the manner in which they interpret diverging performance measures, we also highlight how these two perspectives differ from an "unbiased" view. This alternative perspective highlights the possibility that, when confronted with diverging performance indicators, individuals combine multiple measures into a single measure of performance using weights that reflect the importance of each performance indicator and are not changed depending on whether the performance is high or low.

How decision makers respond to diverging performance measures

Behavioral theory of the firm and control theory

Both the behavioral theory of the firm (Cyert & March, 1963; Greve, 2003a) and control theory (Carver & Scheier, 1981, 1998; Klein, 1989; Lord & Levy, 1994; Miller et al., 1960; Powers, 1973; Vancouver, 2005) seek to explain how performance feedback influences behavior. The behavioral theory of the firm focuses in particular on how performance influences variations in a firm's propensity to change and to innovate. Control theory examines the effect of feedback on a broad range of individual behaviors as well as cognitive and affective processes. Despite the differences in the domains to which these theories are applied, and despite the fact that the findings from these two literatures are rarely integrated, these theories seem to converge in their predictions of how individuals generally respond to the ambiguity introduced by diverging performance measures.

The likely explanation for this striking convergence is that both theories make assumptions rooted in the cybernetic model of individual behavior (Ashby, 1961; Weiner, 1948). For instance, both theories assume that individuals are motivated to act by the desire to reduce negative discrepancies between current outcomes and desired outcomes. In the behavioral theory of the firm (Cyert & March, 1963, pp. 120–122), individuals are said to engage in problemistic searches—searches for solutions to problems evidenced by gaps between the performance and the reference point. Similarly, in control theory (Carver & Scheier, 1998, p. 12), the word “control” refers to “the process of maintaining conformity of a sensed input to a reference value.” Moreover, both theories assume that individuals do not usually allocate their attention equally to all of the issues they confront. Given that individuals have limited attention capabilities, it is argued that greater attention is directed to issues that are more salient whereas attention to issues that are less salient is, at least temporarily, inhibited.

Building on these assumptions, both theories explicitly deal with the question of how individuals respond to multiple and potentially diverging performance measures. They suggest that performance measures are usually attended to in sequence and that the relative importance of performance measures is one of the factors influencing which performance measures are attended to first. Performance measures that are more important are perceived as more salient and therefore are the primary focus of attention. Klein (1989, p. 163), for example, states that “if a salesperson perceives equal discrepancies in goals for completing paperwork and making follow-up calls, and if follow-up calls are viewed as more important, the salesperson will focus his or her attention on calls rather than paperwork.” According to these theories, inhibitory mechanisms limit the recall and retention of details about performance measures that are lower in the hierarchy of importance (Lord & Levy, 1994). However, when the discrepancy between the performance measure highest in the hierarchy of importance and the reference point is eliminated, attention is generally directed to discrepant performance measures lower in the hierarchy. Greve (2003a, p. 72) makes this argument when he notes that, once the most important goals are attained, “managers shift attention among less important goals depending on which goal is in danger of not being met.” Similarly, Carver and Scheier (1998, p. 60) note that:

“In the process of behaving, something being wrong is more likely to draw your attention than something being right, because (in general) when something’s right, you simply move on to the next thing. In contrast when things aren’t right, something else has to be done, adjustments have to be made, before you can move on.”

To our knowledge, no studies examine whether these insights drawn from the behavioral theory of the firm and control theory (hereafter the cybernetic perspectives) correctly predict how individuals respond to the ambiguity introduced by diverging performance measures. In this paper, we take a step toward addressing this gap by examining decision makers confronted with two performance indicators of unequal importance—a primary performance measure (i.e. the most important indicator of overall performance) and a secondary performance measure (i.e., a less important indicator of overall performance). The cybernetic perspectives make two predictions. The first, consistent with well established effects in the literature on change (e.g., Audia et al., 2000; Boeker, 1997; Lant et al., 1992), is that high values of a primary performance measure will decrease decision makers’ propensity to make changes, whereas low values of a primary performance measure will increase it.

Hypothesis 1. When a primary performance measure is above the reference point, decision makers will make fewer changes than when a primary performance measure is below the reference point.

The second prediction is that individuals respond differently to the ambiguity introduced by a secondary performance measure depending on whether the primary performance measure is low or high. When a primary performance measure is low, individuals may be inclined to overlook the ambiguity introduced by a secondary performance measure that exceeds the reference point. They will give limited attention to a secondary performance measure because their focus of attention is on a primary performance measure. Individuals therefore will seek to eliminate the shortfall on a primary performance measure by making changes, and their propensity to change will not be significantly altered by whether a secondary performance measure is high or low. On the other hand, when a primary performance measure is high, greater attention will be given to a secondary performance measure and, as a result, individuals’ decisions will be significantly affected by whether a secondary performance measure is high or low. Individuals will respond to the ambiguity introduced by a secondary performance measure that falls below the reference point by making more changes than if a secondary performance measure was high. In essence, the cybernetic perspectives make the following predictions:

Hypothesis 2.

- (a): When a primary performance measure is above the reference point, decision makers will make significantly more changes if a secondary performance measure is below the reference point than if it is above the reference point.

- (b): When a primary performance measure is below the reference point, decision makers' propensity to change will not be significantly affected by whether a secondary performance measure is below the reference point or above the reference point.

Self-enhancement perspective

The cybernetic perspectives assume that individuals are motivated by the desire to eliminate negative discrepancies between desired outcomes and current outcomes. In contrast, the self-enhancement perspective holds that "people are motivated by the desire to elevate the positivity of their self-conceptions and to protect their self-concepts from negative information" (Sedikides & Strube, 1997, p. 212). The literature indicates that this motivational orientation influences how individuals process evaluative information.

In organizational contexts in particular, numerous studies show that people interpret performance outcomes in a self-enhancing fashion. Individuals, for instance, have been shown to exaggerate favorable evaluations. Harris and Schaubroeck (1988), in a meta analysis of performance evaluation studies, found that self-ratings in virtually all studies they reviewed had higher means than superiors' ratings. Similarly, Clapham (1998) found that management assessment center candidates rated themselves significantly more favorably than did assessors on every one of 16 assessed criteria. Furthermore, in a study of letters to shareholders, Short and Palmer (2003) found that CEOs drew attention to positive performance by making favorable comparisons to competitors and past accomplishments.

Individuals have also been shown to self-enhance by interpreting the meaning of low performance in ways that protect their image as competent managers. A widely reported self-enhancing strategy is to avoid the blame for low performance by attributing it to external and uncontrollable causes. Bettman and Weitz (1983) and Staw, McKechnie, and Puffer (1983), for example, found evidence of this self-enhancing strategy in letters to shareholders. Elsbach and Kramer (1996) evaluated how members of the "top-20" business schools reacted to Business Week's rankings and found that members of business schools that received lower-than-expected rankings often rejected the relevance of the unfavorable ranking by emphasizing favorable performance dimensions.

The literature on self-enhancement, however, suggests that people do not consistently engage in self-serving assessments of reality. Self-enhancing interpretations appear to be more common under conditions of ambiguity, when information can be more easily distorted and idiosyncratic interpretations may be perceived as defensible. For instance, Huber (1991) found that subor-

dinates and superiors had substantial agreement on objective aspects of the subordinate's job, but that subordinates provided self-enhancing ratings on factors that were less verifiable. Similarly, Farh and Dobbins (1989) found that students rated themselves more favorably on an ambiguous performance dimension (participation) than on an unambiguous one (attendance). Moreover, Dunning et al. (1989) found that to the extent that a trait was defined in an ambiguous manner, that is, to the extent that it described a variety of behaviors, individuals provided more self-serving assessments, giving themselves high ranks on positive and low ranks on negative characteristics.

These insights about how people respond to evaluative information suggest that individuals may accurately assess low performance when performance information is unambiguous because in such situations self-serving interpretations would be difficult to defend. However, when performance is ambiguous and individuals are given the opportunity to construct defensible idiosyncratic interpretations, they may assess information in self-enhancing ways, exaggerating favorable information and discounting unfavorable information. If high performance decreases change whereas low performance increases it, as has been demonstrated previously (Greve, 2003a), these self-enhancing interpretations of performance information should impact individuals' propensity to make changes.

Consider again the case in which decision makers are confronted with two performance indicators of unequal importance that may or may not diverge. The literature on self-enhancement does not alter the prediction regarding the effect of a primary performance measure on the decision to change (Hypothesis 1). If individuals are confronted with a single performance measure indicating either high or low performance, there is no ambiguity and therefore there is little room for self-enhancing interpretations. High performance will lead individuals to make fewer changes than low performance. However, the self-enhancement perspective suggests that individuals respond to the ambiguity introduced by a secondary performance measure in a manner opposite to that predicted by the cybernetic perspectives. When a primary performance measure is high, the ambiguity introduced by a secondary performance measure that falls below the reference point will have little effect on the propensity to change. Decision makers will maintain their focus of attention on the primary performance measure because it makes them look competent. They will decrease their propensity to change, and this propensity will not be significantly affected by whether the secondary performance measure is high or low. In contrast, when the primary performance measure is low, the ambiguity introduced by a secondary performance measure that exceeds the reference point will open the door to self-enhancing interpretations motivated by the desire

to maintain an image of competence. Decision makers will shift their attention from the primary performance measure to the secondary one and, consequently, will be significantly less inclined to make changes than decision makers confronted with both a primary and a secondary performance measure falling below the reference point. The self-enhancement perspective therefore suggests the following predictions:

Hypothesis 3.

- (a): When a primary performance measure is below the reference point, decision makers will make significantly fewer changes if a secondary performance measure is above the reference point than if it is below the reference point.
- (b): When a primary performance measure is above the reference point, decision makers' propensity to change will not be significantly affected by whether a secondary performance measure is below the reference point or above the reference point.

Another difference between the cybernetic perspectives and the self-enhancement perspective lies in whether decision makers are accurate in their assessment of the hierarchy among performance measures. The cybernetic perspectives rest on the assumption that decision makers maintain accurate perceptions of the hierarchy among performance measures and that they shift their attention down the hierarchy when performance measures at the top of the hierarchy meet the desired levels. Studies of self-enhancement, instead, suggest that decision makers may rearrange the hierarchy under conditions of ambiguity, when such distortion helps them maintain an image of competence (Elsbach & Kramer, 1996). Under the self-enhancement perspective, when decision makers are confronted with a primary and a secondary performance measure, the propensity to distort information should be greatest when the primary performance measure is low and the secondary performance measure is high. In this scenario, decision makers may preserve a perception of positive performance by increasing the importance they give to the secondary performance measure.

Hypothesis 4. When a primary performance measure is low and a secondary performance measure is high, decision makers are more likely to increase the importance they give to the secondary performance measure than when any of the other three performance combinations occurs.

An alternative view

These two perspectives differ from how “unbiased” decision makers would behave. If decision makers are

unbiased in the manner in which they respond to diverging performance indicators, the impact of a secondary performance measure on the decision to change should not be altered by whether a primary performance measure is high or low. Moreover, decision makers would not change the importance assigned to a secondary performance measure depending on the situation they are confronting.

Study 1

In Study 1, we examined the impact of multiple performance indicators using an experimental design. Participants, playing the role of a manager, were given both a primary and a secondary performance measure and were asked to make relevant organizational changes. The experimental approach enabled us to determine how participants respond to the ambiguity introduced by diverging performance measures.

Method

Participants and design

Eighty-nine undergraduates from a west-coast university, 46 women and 43 men, participated in the study in partial fulfillment of course credit. The experimental design consisted of a 2 (Primary Measure: Low vs. High Performance) × 2 (Secondary Measure: Low vs. High Performance) between-subjects factorial design.

Materials and procedure

All materials for the study were presented on paper. Instructions on the first page informed participants that the study consisted of three parts: first, they would read a general description of the radio industry; second, they would be given information about a specific company within the radio industry named Norfolk Radio; and finally, they would decide whether to change the programming schedule. Participants were told that they would be playing the role of the general manager of Norfolk Radio. This manipulation was intended to ensure that participants internalized the performance of the company.

The background information on the radio industry consisted of a brief description of the use of “formats,” which were defined as “a combination of program content, announcer style, timing of program and commercial material, and methods for listener feedback and quality control.” Participants were then given several examples of formats. The background information also informed participants that radio stations use two measures to assess performance: *Share* and *Time Spent Listening*. *Share* was defined as the average proportion of all listeners over 12 years old tuned in to a station during the broadcast week. *Time*

Spent Listening was defined as the average time listeners spent tuned in to the station. Reference points separating high performance from low performance for both measures were set by informing participants that (a) “a radio station must get a Share of at least 5% to attract significant revenues,” and (b) “the industry norm is a weekly average Time Spent Listening of 10 hours.” These values were determined using information published by Arbitron, the leading audience measurement firm. Interestingly, although 10 h of time spent listening to the radio might seem too high, Arbitron (2002) reports that in 2001 people spent on average more than 20 h per week listening to the radio. We chose to use 10 h as the reference point, rather than 20 h, because it allowed us to manipulate Time Spent Listening, setting high and low values that participants were likely to perceive as realistic. For the precise values, see Experimental Manipulation.

The two performance measures were also explicitly given a hierarchy, with Share as the primary measure, and Time Spent Listening as the secondary measure. The hierarchy reflects the importance radio station managers and advertisers attribute to these performance indicators (Arbitron, 2002). Participants were explicitly told that Share was “the most critical indicator of a radio station’s performance,” while “industry analysts and advertisers depended less on Time Spent Listening.” Participants then read a brief description of Norfolk Radio, which included information about the company founding and examples of current radio programming. Participants were also given Norfolk Radio’s 1998 and 2002 performance values for both Share and Time Spent Listening.

The third section instructed participants that as Norfolk Radio’s general manager, they needed to determine whether current performance warranted a change to the current programming schedule. Participants were reminded that format changes are costly for the station, but may have a positive impact on Norfolk Radio’s performance. Participants then indicated the percentage of current programming hours that they would change to allow for the introduction of different formats, ranging from 0 to 100% change. Following this decision, participants were instructed to write a brief essay explaining their decision.

Participants then responded to a series of questions assessing the impact that Share and Time Spent Listening had on their decision. These questions were followed by a set of manipulation checks, aimed at measuring whether participants correctly recalled the actual performance values of Share and Time Spent Listening.

Experimental manipulation

Participants were randomly assigned conditions in which they received different values of the two performance outcomes. In the *Low-Share* conditions, partic-

ipants were informed that Share dropped from 5% in 1998 to 1% in 2002. In the *High-Share* conditions, participants were informed that Share had increased from 5% in 1998 to 9% in 2002. In the *Low-Time Spent Listening* conditions, participants were told that Time Spent Listening decreased from 10 h per week in 1998 to 2 h per week in 2002. In the *High-Time Spent Listening* conditions, participants were told that Time Spent Listening increased from 10 h per week in 1998 to 18 hours per week in 2002.

Results

Manipulation checks

To determine whether participants correctly distinguished the value of Share and Time Spent Listening in the Low and High conditions, participants recalled the value of each outcome. A main effect for Share indicates that participants correctly distinguished the value of Share in the *Low-Share/Low-Time Spent Listening* (LL) and *Low-Share/High-Time Spent Listening* (LH) vs. *High-Share/Low-Time Spent Listening* (HL) and *High-Share/High-Time Spent Listening* (HH) conditions, $M_{Low} = 1.79$ vs. $M_{High} = 8.98$, $F(1, 88) = 461.39$, $p < .001$.

A main effect for Time Spent Listening indicates that participants correctly distinguished the value of Time Spent Listening in the LL, HL vs. LH, HH conditions, $M_{Low} = 2.60$ vs. $M_{High} = 17.60$, $F(1, 86) = 562.16$, $p < .001$. Note that two participants did not complete the Time Spent Listening manipulation check.

We also sought to confirm that participants correctly encoded the hierarchical relationship between the two performance measures. Participants were asked to identify which criterion industry analysts and advertisers believed was more important and were given three choices: Share, Time Spent Listening, and Same. Most participants ($M = 74.16\%$) considered Share as the most important objective performance measure, versus Time Spent Listening ($M = 15.73\%$) and Same ($M = 10.11\%$), $\chi^2(2, n = 89) = 67.17$, $p < .001$.¹ There were no significant effects of Share or Time Spent Listening on objective importance, indicating that assessments of objective importance did not vary by condition. Furthermore, there were no other effects of manipulation check variables.

Tests of the hypotheses

A 2 (Share: low vs. high) \times 2 (Time Spent Listening: low vs. high) analysis of variance was conducted on

¹ As a precautionary measure, we ran the remainder of the analyses presented below after removing the 26% of participants not correctly responding to the manipulation check (i.e., responding either TSL or Same). Results from these analyses reveal identical patterns to results with the entire sample.

participants' propensity to change.² A significant main effect of Share was observed, $F(1,88) = 22.00$, $p < .01$, as predicted by Hypothesis 1: participants in the Low Share conditions changed the programming schedule significantly more than participants in the High Share conditions, $M_{\text{Low}} = 47.21$ vs. $M_{\text{High}} = 26.02$. Also, a main effect for Time Spent Listening was found, indicating that participants in the Low Time Spent Listening conditions changed the programming schedule significantly more than participants in the High Time Spent Listening conditions, $M_{\text{Low}} = 42.69$ vs. $M_{\text{High}} = 29.20$, $F(1,88) = 9.42$, $p < .01$. Although the traditional analysis of variance does not capture an interaction between Share and Time Spent Listening,³ $F(1,88) = 0.82$, *n.s.*, simple effects analyses revealed that the effect of Time Spent Listening on change was significant only when Share was low, $M_{\text{LL}} = 56.24$ vs. $M_{\text{LH}} = 38.19$, $F(1,88) = 7.45$, $p < .01$. Time Spent Listening did not significantly influence change decisions when Share was high, $M_{\text{HL}} = 30.83$ vs. $M_{\text{HH}} = 21.00$, $F(1,88) = 2.50$, *n.s.* (see Table 1, Fig. 1). These results therefore support Hypothesis 3 and contradict Hypothesis 2. Participants responded to the ambiguity introduced by a secondary performance measure in a self-enhancing manner, taking that measure into account in their decision to change only when it enhanced the overall performance.

Hypothesis 4 predicts that decision makers confronted with a low primary performance measure and a high secondary performance measure are more likely to re-arrange the hierarchy of the performance measures, giving greater importance to a secondary performance measure, than decision makers in other conditions. While the manipulation check mentioned above demonstrated that most participants correctly recalled Share as the most important "objective" measure of performance, we also asked which criterion was more important in their assessment. They were given three options: Share, Time Spent Listening, or Same. Of course, this item captures only cases in which participants were aware and admitted that they rearranged the importance of the

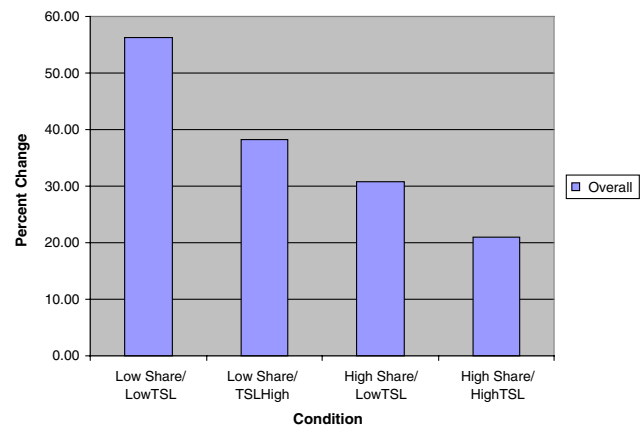


Fig. 1. Mean percent change by condition—Study 1.

two performance measures, thus understating possible effects. It may be that individuals changed the hierarchy of the two performance measures but were unaware of the change or unwilling to admit that the change occurred. Tabulating the frequencies of personal importance to objective importance showed that forty-nine participants changed the hierarchy of the two performance measures. Of these 49, 29 considered Time Spent Listening as the most important measure and 24 considered Time Spent Listening as important as Share. We then sought to determine under what conditions such a shift was more likely to occur. A logistic regression in which the dependent variable was 1 when the importance shift occurred and 0 otherwise did not show a main effect of Share (coeff = 0.349; $z = 1.3$). However, a planned comparison indicated that participants' inclination to increase the importance given to Time Spent Listening was significantly higher when Time Spent Listening was high and Share was low than in the other three conditions $M_{\text{LH}} = 71\%$ vs. $M_{\text{LL, HL, HH}} = 50\%$, coeff = .35; $z = 2.12$. This evidence supports Hypothesis 4 and therefore suggests that participants confronted with diverging performance measures self-enhance by re-arranging the hierarchy of the performance measures in a manner that helps them maintain a sense of positive performance.

Supplemental analysis

To further test for the presence of a self-enhancing response to diverging performance indicators, we examined whether individuals who regard themselves as above-average show a stronger self-enhancing response to diverging performance measures. If, under conditions of ambiguity, individuals discount negative information to protect their self-images, as the literature on self-enhancement suggests, then those who have more positive self-images should show a stronger tendency to self-enhance. Therefore, we expected that high scorers on a dispositional measure of self-enhancement would give greater importance to Time Spent Listening when Time

² A second set of analyses was also performed with an analysis of covariance, controlling for age, gender, and self-reported grade point average. Of the 89 participants, 4 participants who did not complete age, gender, or grade point average questions were removed. The set with controls had 42 women and 42 men. No significant differences were found between analyses with and without controls. Gender, grade point average, and age did not significantly impact any of the findings.

³ The traditional analysis of variance is particularly powerful when used for detecting interactions characterized by completely opposing effects (i.e., cross-over interactions) but may fail to detect other forms of interactions (e.g., Judd & McClelland, 1989; Rosenthal & Rosnow, 1985). Simple effects, that is, planned comparisons between categories of one variable (e.g., High and Low Time Spent Listening) within a single level of another variable (e.g., High Share or Low Share) provide a more appropriate and more powerful test of the interaction patterns predicted in Hypothesis 2 and 3.

Table 1

Means, standard deviations, and *N* for change, by condition and by low vs. high self-enhancement (by median split)—Study 1

	Overall		Low Self-enhancers		High Self-enhancers	
	Mean (<i>SD</i>)	<i>n</i>	Mean (<i>SD</i>)	<i>n</i>	Mean (<i>SD</i>)	<i>n</i>
Low Share/Low TSL	56.24 (23.78)	21	54.56 (22.48)	9	57.50 (25.63)	12
Low Share/High TSL	38.19 (22.16)	21	46.11 (25.71)	9	32.25 (17.96)	12
High Share/Low TSL	30.83 (18.51)	24	33.64 (22.59)	11	28.46 (14.77)	13
High Share/High TSL	21.00 (21.18)	23	19.85 (20.03)	13	22.50 (23.60)	10

Note. TSL stands for Time Spent Listening, the secondary performance measure.

Spent Listening was high and Share was low than low scorers. Several days prior to the study, all participants completed the “How I See Myself Questionnaire” (Taylor, Lerner, Sherman, Sage, & McDowell, 2003), a dispositional measure of self-enhancement. Participants rated themselves in comparison with their peers on 21 positive and 21 negative items. The 7-point scale measured comparisons from 1 (*much less than the average college student of my age and gender*) to 7 (*much more than the average college student of my age and gender*). Negative items were reverse scored, such that high scorers said they rated higher on positive items and lower on negative items than others. The HSM questionnaire had a high degree of reliability ($\alpha = .90$).

To analyze the change decisions made by participants with High and Low dispositional levels of self-enhancement, we assigned participants to high and low levels using a median-split.⁴ The main effect of Share held for both high self-enhancers, $M_{Low} = 44.88$ vs. $M_{High} = 25.87$, $F(1,46) = 10.24$, $p < .01$, and low self-enhancers, $M_{Low} = 50.33$ vs. $M_{High} = 26.17$, $F(1,41) = 11.29$, $p < .01$. The simple effect of Time Spent Listening when Share was low held for high self-enhancers, $M_{LL} = 57.50$ vs. $M_{LH} = 32.25$, $F(1, 46) = 8.94$, $p < .01$, but it was not significant for low self-enhancers, $M_{LL} = 54.56$ vs. $M_{LH} = 46.11$, $F(1,41) = 0.64$, *n.s.* (Table 1, Fig. 2). Furthermore, the simple effect of Time Spent Listening when Share was high was not significant for high self-enhancers, $M_{HL} = 28.46$ vs. $M_{HH} = 22.50$, $F(1,46) = 0.48$, *n.s.*, and it was also not significant for low self-enhancers, $M_{HL} = 33.64$ vs. $M_{HH} = 19.85$, $F(1,41) = 2.25$, *n.s.* Note also that the main effect of self-enhancement on change was not significant, $F(1,87) = 0.41$. These results indicate that the effect predicted in Hypothesis 3 is led by participants with a high disposition to self-enhance, and these results therefore lend additional support to the view that self-enhancement influences how people respond to diverging performance measures. Moreover, these results provide additional evidence that runs counter to Hypothesis 2 because low self-enhancers did not conform to the prediction made by the cybernet-

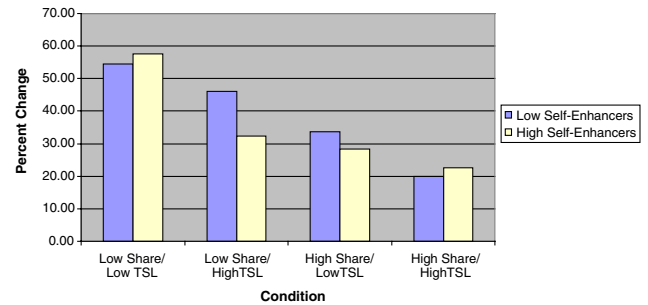


Fig. 2. Mean percent change by condition and by low/high self-enhancement—Study 1.

ic perspectives. The results also contradict the “unbiased” perspective, whereby people base their decision to make changes by assigning weights to performance measures that reflect their relative importance and keeping these weights unchanged across situations.

Overall, the pattern of findings of Study 1 is consistent with a self-enhancing bias. A diverging secondary performance measure affected the decision to change only when it helped individuals to maintain a sense of positive performance, that is, when a primary performance measure was low. Furthermore, individuals were more likely to increase the importance assigned to a secondary performance measure when this measure helped them maintain a sense of positive performance, that is, when a primary performance measure was low and a secondary performance measure was high. Finally, the finding that a secondary performance measure affected the decision to change when a primary performance measure was low held only for individuals who had high scores on a self-enhancement scale.

Study 2

To verify whether self-enhancing responses to diverging performance measures extend to field settings, we tested Hypothesis 1–3 by examining the impact of two key organizational performance measures—revenue growth and profitability—on top executives’ propensity to introduce new products. Like changes to the programming schedule in Study 1, product introductions are an important form of organizational change (e.g.,

⁴ We find the same results when we use a general linear model in which self-enhancement is treated as a continuous variable.

Barnett & Freeman, 2001). When a radio station adds a radio format, adjustments need to be made to roles, personnel, and established routines. Similarly, when an organization introduces a new product, changes in roles, personnel, and routines are necessary especially in the parts of the organization dedicated to production and distribution.

Sample and data collection

Our sample includes information about hard disk drive manufacturers. We chose to focus on this industry for two main reasons. First, this is an industry that has historically had high rates of innovation accompanied by the frequent introduction of new products (Christensen & Bower, 1996). Second, although product introductions data is rare, we had access to a publication, *Disk Trend*, that reported information about the products offered by organizations in this industry in great detail. This publication was available until 1999.

The sample includes all U.S. public firms specializing in hard disk drives. Private firms and divisions of diversified firms were excluded because data about revenues and especially profits was often unavailable for these organizations. We defined firms specializing in hard disk drives as those that derived 75% or more of their revenues from sales of hard disk drives, and we obtained their performance information from Compustat, a financial database listing public firms by industry. The eight firms that met this requirement were: Maxtor, Micropolis, Miniscribe, Priam, Western Digital, Conner Peripherals, Seagate, and Quantum. The sample includes yearly observations beginning with the year in which the firm went public and ending with the cessation of operation or in 1999 for those that remained in existence. There are 88 observations, each corresponding to an organization-year. On average these firms introduced 15 new hard disk drives per year. In addition to these 8 firms, we also report a supplemental analysis of private and diversified public firms for which performance information was available.

A key issue in Study 2 was to determine the hierarchy of the performance measures. Although both revenue growth and profitability are critical to an organization's success, industry accounts suggest that, in this particular setting, revenue growth was clearly regarded as more important than profitability. The industry's phenomenal growth rate was a key reason for this hierarchy. Total industry revenues of U.S. hard disk drive manufacturers went from \$1.7 billion in revenues in 1977 to over \$23 billion in 1999. In this context of continual expansion, success was often unequivocally linked to record growth in revenues. Conner Peripherals, for example, was celebrated as one of the greatest success stories both inside and outside the industry, not because of its profitability, but because it became the fastest-growing manufactur-

ing start-up in American business history and only the second company ever to reach the ranks of the Fortune 500 in its third year of sales (Richards, 1990). Profitability was also highly valued, but tended to be viewed more as an important condition to sustain revenue growth than as a primary goal.

To corroborate the view that revenue growth was viewed as more important than profitability, we examined the correlation between these two performance indicators and share price. While revenue growth and profitability are examples of measures that reflect the results of past activities, share price is a market valuation measure that reflects investors' expectations about future performance (Meyer, 2002). Arguably, if revenue growth was viewed as more important than profitability, then it should be more strongly related to share price than profitability. Information about the share price of U.S. public firms specializing in hard disk drives was obtained from Compustat. We computed the average share price in a given year by averaging the share price at the end of each month. While the correlation between revenue growth and the average share price was positive and significant, $r = .28, p < .05$, we did not find a significant correlation between profitability and the average stock price, $r = .07, n.s.$ Thus this evidence seems consistent with the hypothesized hierarchy of revenue growth and profitability.

Variables

In this study, the dependent variable is the count of products introduced in a particular year. The independent variables are revenue growth and profitability in relation to their respective reference points. To determine the reference points we followed the procedure adopted in studies of the effect of performance on organizational change (Greve, 2003a). This literature usually makes the assumption that the reference point or aspiration level against which organizational members evaluate performance is determined by the organization's recent performance history (historical aspiration level) or by the performance of similar organizations (social aspiration level) (Cyert & March, 1963; Greve, 2003a). In preliminary analyses, we explored the effect of revenue growth and profitability using both reference points. The social aspiration level was the mean value of the other firms in the sample, whereas the historical aspiration level was the organization's performance in the previous year. These analyses showed that the performance measures had a greater effect on the decision to introduce new products when we used the historical aspiration level than when we used the social aspiration level. We interpret this evidence as suggesting that organizational members in these firms paid greater attention to historical aspiration levels. Therefore the results reported here are based on historical aspiration levels.

In preliminary analyses, we also experimented with different lags for the performance measures and found that the decision to introduce new products in year t was influenced most strongly by the performance in year $t - 1$. This is in line with the view that firms tend to have a portfolio of products ready for launch. Product introductions, like innovation launches (Greve, 2003b), tend to reflect decisions to launch products that are internally available rather than decisions to initiate new product development efforts that may take years before a new product is launched.

The analyses also include several control variables. We include organizational size measured as the log of a firm's assets and the number of products offered by the firm, because larger organizations and organizations that have a larger product portfolio may be more likely to introduce new products. We also include the number of new products introduced by competitors in the previous year, because firms may introduce more products under the influence of what other firms do. Finally, we include the calendar year in order to pick up industry-level processes linked to the passage of time that may influence product introductions. Like the performance variables, these control variables correspond to the year preceding product introductions.

Model

The dependent variable is a count variable that takes only non-negative values. Because the distribution of count variables is not normal and because ordinary linear regression assumes that errors are normally distributed, we use negative binomial regression, which is a special instance of the poisson regression. Poisson regression takes into account the special nature of count data (e.g., truncation to zero, right skewed distribution). The choice between poisson regression and negative binomial regression is dictated by whether the variance of the dependent variable exceeds the mean (Cameron & Trivedi, 1998). Because we found strong evidence of overdispersion (likelihood-ratio test = 643.3, $p < .01$), we report negative binomial regressions. To control for unobserved organization-specific effects that may be a source of serial correlation, we estimate negative

binomial regressions with random effects (Hausman, Hall, & Griliches, 1984). The random effect specification includes an organization-specific effect that permits observations of the same organization to be correlated across periods:

$$\lambda_{it} = \exp(\mathbf{X}_{it}\beta + \mu_o + \mu_i),$$

where λ_{it} is the number of new products introduced by organization i in year t , \mathbf{X}_{it} is the vector of the covariates for organization i in year t , β is the vector of the coefficients to be estimated, μ_o is the overall intercept, and μ_i is the organization-specific effect.

Results

Table 2 reports descriptive statistics and correlations among the variables, whereas Table 3 reports the results of the negative binomial regressions. A negative coefficient indicates a negative effect of an independent variable on the probability of a product introduction, whereas a positive coefficient indicates a positive effect. Because the negative binomial regression models the natural log of the dependent variable as a linear function of the coefficients, it is necessary to exponentiate coefficients times the change in the independent variable to compute the magnitude of these effects (e.g., Cameron & Trivedi, 1998). The only control variable that shows consistent effects is organizational size. The coefficient indicates that larger firms are more likely to introduce new products. Model 2 shows that revenue growth has a negative effect on the rate of product introductions, as suggested in Hypothesis 1 and consistent with past studies that focused on a single performance indicator (e.g., Audia et al., 2000; Boeker, 1997; Lant et al., 1992). The coefficient implies that a one-standard-deviation increase in revenue growth is associated with a 15.1% decline in the rate of product introductions ($e^{(0.691 * -0.238)} - 1 = -0.151$). In Model 3, we examine the effect of profitability within high and low levels of revenue growth by adding the same contrasts used in Study 1. The results lend support to the self-enhancement perspective: LL_LH is negative and significant whereas HH_HL is negative but not significant. Organizations' propensity to introduce new products is affected

Table 2
Descriptive statistics and correlations—Study 2

	Mean	SD	Minimum	Maximum	1.	2.	3.	4.	5.	6.
1. Product introductions	15.19	15.56	0	76						
2. Log of organizational size	6.09	1.17	3.77	8.81	0.646					
3. Number of products	26.47	23.26	5	116	0.568	0.675				
4. Product introductions in the industry	83.55	46.50	4	156	0.322	0.646	0.358			
5. Calendar year	1990.47	4.27	1982	1998	0.445	0.744	0.377	0.846		
6. Revenue growth minus the reference point	-0.08	0.69	-3.00	2.65	-0.137	0.011	0.021	0.104	0.002	
7. Profitability minus the reference point	-0.00	0.17	-0.84	0.69	-0.039	0.018	-0.043	-0.125	-0.067	0.097

Note. Correlations with absolute values greater than .3 are significant at the .05 level; $N = 88$.

Table 3
Negative binomial regressions of product introductions—Sample including U.S. public firms specializing in hard disk drives (N = 88)—Study 2

	1	2	3	4	5	6
Log of organizational size	0.428*** (0.14)	0.422*** (0.137)	0.482*** (0.135)	0.469*** (0.144)	0.491*** (0.192)	0.489*** (0.144)
Number of products	-0.001 (0.004)	-0.001 (0.003)	-0.001 (0.004)	-0.001 (0.004)	-0.002 (0.004)	-0.002 (0.004)
Product introductions in the industry	-0.003 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.003 (-0.002)	-0.002 (0.002)	-0.001 (0.002)
Calendar Year	0.057 (0.037)	0.043 (0.037)	0.034 (0.035)	-0.051 (0.037)	0.047 (0.036)	0.026 (0.039)
Revenue growth minus the reference point		-0.238*** (0.096)	-0.249*** (0.098)			
LL vs. LH			-0.178*** (0.087)		-0.174 (0.11)	
HH vs. HL			-0.04 (0.118)		0.068 (0.121)	
Profitability minus the reference point				-0.508 (.438)		
Revenue growth above the reference point						-0.576** (0.234)
Revenue growth below the reference point						-0.035 (0.162)
Constant	-115.31 (73.86)	-89.07 (73.58)	-71.28 (70.26)	-104.93 (73.12)	-96.61 (71.32)	-53.91 (77.04)
Log likelihood	-300.38	-297.41	-295.25	-299.72	298.39	-295.75

Note. LL = Low revenue growth and Low profitability; LH = Low revenue growth and High profitability; HH = High revenue growth and High profitability; HL = High revenue growth and Low profitability.

*p < .10.
** p < .05.
*** p < .01.

by profitability but only when revenue growth is below the reference point, as predicted in Hypothesis 3. Focusing on organizations experiencing low revenue growth, the coefficient for the contrast LL vs. LH implies that those with profitability above the reference point are 35% less likely to introduce new products than those with profitability below the reference point $e^{(1*-0.178)} - e^{(-1*-0.178)} = -0.357$.

To test the robustness of our results, we conducted two additional sets of analyses. First, if, in this particular setting, revenue growth is indeed a more important performance measure than profitability and if decision makers respond to profitability depending on whether revenue growth is above or below the reference point, we should find a weaker pattern of results when we reverse the hierarchy of the two performance measures. As the results in Models 4 and 5 show, profitability does not have a significant negative effect on the rate of product introductions and the LL vs. LH contrast is no longer significant. Second, we sought to corroborate the results of the models restricted to U.S. public firms specializing in hard disk drives by also analyzing private firms and divisions of diversified firms such as HP and IBM for which revenues and profit data were reported in *Disk Trend*. It must be noted that the profit data reported by private firms and divisions of diversified firms tends to be more subject to idiosyncratic calculations than the profit data reported by public organizations. This sample includes 201 organization-year observations of 26 organizations. Note that organizational size was measured as the log of revenues in these models and that we added the percentage of revenues derived from transactions within the organization (i.e., captive sales) as a control. The results in Table 4 indicate that we find virtually the same pattern of results supporting Hypothesis 1 and 3 when we use this broader sample.

Discussion

Our experimental and field studies indicate that people charged with deciding whether to make changes respond to diverging performance indicators in a self-enhancing manner. Decision makers gave importance to a secondary performance indicator only when it helped them maintain a sense of positive performance, that is, when a secondary performance measure was high and a primary performance measure was low. Our findings contradict the hypothesis made by the behavioral theory of the firm and control theory that decision makers give importance to a secondary performance measure only when it is low and a primary performance measure is high. Moreover, our findings contradict the view that decision makers respond to diverging performance measures in an unbiased manner by giving the same impor-

Table 4
Negative binomial regressions of product introductions—Sample including U.S. public firms, private firms, and divisions of diversified firms (N = 201)—Study 2

	1	2	3	4	5	6
Log of organizational size	0.247*** (0.084)	0.252* (0.083)	0.259*** (0.081)	0.256* (0.083)	0.264*** (0.081)	0.248*** (0.083)
Number of products	0.003 (0.003)	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)	0.002 (0.002)	0.002 (0.003)
Product introductions in the industry	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (-0.001)	-0.001 (0.001)	-0.001 (0.001)
Percentage of internal sales	-0.666*** (0.245)	-0.681*** (0.251)	-0.681*** (0.243)	-0.666*** (0.245)	-0.671*** (0.24)	-0.712*** (0.256)
Calendar Year	0.047* (0.024)	0.038 (0.023)	0.045* (0.023)	0.049** (0.024)	0.056** (0.023)	0.036 (0.023)
Revenue growth minus the reference point		-0.269*** (0.083)	-0.269*** (.083)			
LL vs. LH			-.139** (.064)			
HH vs. HL			-.006 (.083)			
Profitability minus the reference point				0.015 (0.008)		
Revenue growth above the reference point						-0.471*** (0.17)
Revenue growth below the reference point						-0.091 (0.151)
Constant	-95.32*** (47.97)	-77.72* (46.56)	-90.41** (45.62)	-99.72** (47.36)	-113.41** (46.61)	-73.39 (47.25)
Log likelihood	-600.95	-595.58	-593.23	-599.53	-596.92	594.43

* p < .10.
** p < .05.
*** p < .01.

tance to a secondary performance measure across situations, irrespective of whether a primary performance measure is high or low.

Our findings suggest that the impact of diverging performance measures on individuals' decisions to change should result in a modification to the well-accepted prediction that high performance reduces change whereas low performance increases it. Our results indicate that when a primary performance measure is high, decision makers' propensity to change is not affected by a diverging secondary performance measure. This presumably occurs because decision makers give little attention to a secondary performance indicator exhibiting poor performance. However, when a primary performance measure is low, decision makers change significantly less than the usual prediction of increased change driven by low performance suggests, because they give greater importance to a secondary performance measure that indicates a favorable outcome. In other words, under conditions of ambiguity deriving from diverging performance measures, low performance increases decision makers' propensity to change less than previously thought.

Although our empirical analyses focused on two performance indicators, the literature on self-enhancement (Dunning et al., 1989; Johns, 1999) implies that self-enhancing interpretations of low performance should occur, or be even stronger, when the number of performance measures under consideration is greater, because a greater number of performance measures implies greater ambiguity surrounding the precise meaning of the overall performance. Ambiguity, in turn, makes it easier for decision makers motivated to maintain an image of competence to re-evaluate the importance of positive performance indicators in a self-enhancing manner. Indeed, as the opening quote by Eisner suggests, decision makers may take advantage of the existence of numerous performance indicators to obscure performance on primary indicators.

If the ambiguity introduced by diverging performance measures weakens the effect of low performance on decision makers' propensity to make changes, we should expect less change in response to low performance than is implied by the often-held view that failure instigates change, especially in real organizations that tend to have multiple and weakly linked performance indicators. The empirical evidence from the few organization-level studies that separate the effect of performance above and below the reference point is consistent with this suggestion. For instance, a study of radio stations shows that audience share below the reference point increased the probability of adopting new radio formats less strongly than audience share above the reference point decreased it (Greve, 1998). In other words, failure on the key performance indicator for radio stations increased change significantly less than success on that same performance indicator

decreased change. Similarly, a study of shipbuilding firms (Greve, 2003b) shows that return on assets above the reference point decreased managers' propensity to launch innovations, whereas return on assets below the reference point did not have any effect. Interestingly, supplemental analyses in Study 2 reveal the same pattern in our data. When we separate revenue growth above and below the reference point (Table 3, Model 6; Table 4, Model 6), we find that revenue growth above the reference point has a negative and significant coefficient whereas revenue growth below the reference point is not significant. Although our results suggest that a self-enhancing response to diverging performance measures is a reason for the weaker effect of low performance on change, it must be noted that other processes might contribute to this kink in the performance–change curve. For instance, decision makers may fail to make changes following the receipt of negative performance because their efforts are stifled by the presence of bureaucratic constraints (Greve, 1998). An interesting direction for future studies would be to examine the extent to which these different processes contribute to decision makers' reluctance to change in response to low performance.

More broadly, this paper suggests that research regarding the effect of performance on the decision to change may benefit from a closer integration with the literature on self-enhancement. A fruitful approach for future research may be to investigate factors, both dispositional and situational, that influence the tendency to engage in self-enhancing interpretations of ambiguous performance information. For instance, the supplemental analysis in Study 1 suggests that an important dispositional factor is the extent to which people regard themselves as better than their peers. Another potentially important dispositional factor is whether people conceive ability as a malleable quality or a fixed entity. Dweck and Leggett (1988) suggest that individuals who construe ability as a malleable quality that can be continually enhanced tend to regard deficient performance as an opportunity for learning and future advancement. On the other hand, individuals who construe ability as a fixed entity tend to regard deficient performance as a threat to their self-image. In addition, Wood and Bandura (1989) demonstrate that these conceptual frames can be primed by social contexts, implying that organizational contexts may induce different conceptions of ability. Past studies do not examine how individuals holding these different conceptions interpret ambiguous performance information, but it may be that those holding a conception of ability as a fixed entity will be more likely than those holding a conception of ability as a malleable quality to interpret ambiguous information in a self-enhancing way.

Situational moderators may also impact the interpretation of diverging performance outcomes. For

instance, the type of scrutiny defined as “the monitoring of behavior or performance, frequently with the goal of increasing accountability” (Johns, 1999, p. 19) is also likely to influence individuals' propensity to self-enhance. Proactive scrutiny, which occurs when decision makers anticipate being scrutinized before they act, is likely to temper self-enhancement because it stimulates self-evaluation that contributes to more realistic thinking. In contrast, retroactive scrutiny, which occurs after decision makers act and typically follows some unforeseen negative consequences, likely exacerbates self-enhancement because it heightens the motivation to defend one's behavior. In addition, the distribution of power may also influence how organizational members respond to ambiguous performance information. Ocasio (1995) argued that deficient organizational performance leads to fewer changes to the power distribution when power is concentrated in cohesive top management groups than when it is fragmented across multiple constituencies. It may be that the degree of power concentration also influences how ambiguous performance information is interpreted. Self-enhancing interpretations may be more pronounced when power is concentrated than when it is fragmented. In the latter case, divergent interests among those who hold power may favor mixed interpretations of ambiguous performance.

Insights from research of self-enhancement might also help researchers interested in the impact of performance on change to understand how people choose among multiple reference points, a major unanswered question in studies of decision making (Kahneman, 1992). Reference points demarcate the line between success and failure, but research on self-enhancement suggests that individuals may choose reference points in ways that preserve perceptions of success. Consider, for example, how people form social aspiration levels. Social psychologists show that people often maintain a positive self-image by choosing less-fortunate others as comparisons (Taylor & Lobel, 1989; Wills, 1981) and also by distancing themselves from similar individuals who outperform them (Tesser, 1988). Similarly, organizational researchers find evidence of self-enhancement in how boards choose comparison companies when justifying the compensation of CEOs (Porac, Wade, & Pollock, 1999). While boards generally justify CEO compensation by comparing their firms to other companies within the same industry, they tend to choose outside of their own industry when their firms perform poorly, their industry performs well, or CEOs are highly paid. Moreover, the comparisons outside of the primary industry tend to be comparisons to lower-performing firms. While these studies suggest that self-enhancement influences the choice of comparable others, researchers could carry this insight into

investigations of the effect of performance on change. We speculate that to the extent that the tendency to self-enhance influences the choice of reference points, perceptions of success and the associated reluctance to change an existing strategy might be more pervasive than is often thought.

Our results also add to research on self-enhancement and in particular to its link to risk taking. Past studies show that correlates of the disposition to self-enhance such as self-esteem and unrealistic optimism increase risk taking (Baumeister, Heatherton, & Tice, 1993; Josephs, Larrick, Steele, & Nisbett, 1992; McFarlin & Blascovich, 1981). Presumably people who think highly of themselves take greater risks because they think that they are less likely to suffer harm than their peers (Burger & Burns, 1988; Taylor & Lobel, 1989). However, if we equate making changes with taking greater risks, the results of Study 1 identify a particular context in which the motivation to protect one's positive image appears to interfere with risk taking. While low performance generally increases risk taking, self-enhancers may be less likely to take risks than non-self-enhancers because they are motivated to give greater attention to favorable performance outcomes that call for less risky actions.

Limitations

Although a strength of this paper is that it investigates the same relationships in both an experimental and a field study, the parallel between the two studies is far from perfect. While in the experimental study we were able to provide clear instructions indicating which performance measure was most critical and which performance measure was less important, in the field study we assumed that revenue growth was the most critical performance measure and profitability was less important. Although it is reassuring that revenue growth was more strongly linked to the share price than profitability and that the pattern of results supports the hypothesized hierarchy of the two performance measures, this difference between the two studies should be kept in mind. Furthermore, while the experimental study examined decisions made by individuals, the field study focused on the introduction of new products, which are often decisions made by groups of executives. This difference, however, should not limit the validity of our conclusions, because research has shown similar self-enhancing tendencies at the group level (Johns, 1999; Polzer, Kramer, & Neale, 1997).

Our findings may be also limited by specific features of the decision making context. In the experimental study, participants playing the role of a manager made decisions about changing the programming schedule. Although our results suggest that this procedure is suffi-

cient to ensure that participants internalize the performance of the company, the changes made remain hypothetical, and participants are somewhat removed from their consequences. Future research could ensure that similar self-enhancing decisions also occur in contexts in which decision makers face consequences. Moreover, participants were given performance information, and no search process for such performance information was necessary. Future work should investigate decision makers' search processes in contexts where information is not so readily available and reliable. Future research could also examine decision makers in industries in which more than two performance indicators are commonly used. This work could shed additional light on the hierarchies that managers create to interpret performance outcomes.

Conclusion

In closing, we have found that people react to diverging performance measures in a self-enhancing manner. Evidence from both an experimental and a field study suggests that decision makers are motivated to maintain a sense of positive performance and actively rearrange the hierarchy of diverging performance measures to reach this aim. Furthermore, the tendency to self-enhance makes decision makers reluctant to acknowledge the need for change.

References

- Arbitron (2002). *Radio's leading indicator*. Washington, DC: Arbitron.
- Ashby, W. R. (1961). *An introduction to cybernetics*. London: Chapman.
- Audia, P. G., Locke, E. A., & Smith, K. G. (2000). The paradox of success: An archival and laboratory study of strategic persistence following radical environmental change. *Academy of Management Journal*, 43(5), 837–854.
- Barnett, W. P., & Freeman, J. (2001). Too much of a good thing? Product proliferation and organizational failure. *Organization Science*, 12, 539–558.
- Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1993). When ego threats lead to self-regulation failure: Negative consequences of high self-esteem. *Journal of Personality and Social Psychology*, 63, 989–1003.
- Bettman, J., & Weitz, B. (1983). Attributions in the boardroom: Causal reasoning in corporate annual reports. *Administrative Science Quarterly*, 28, 165–183.
- Boeker, W. (1997). Strategic change: The influence of managerial characteristics and organizational growth. *Academy of Management Journal*, 40, 152–170.
- Burger, J. M., & Burns, L. (1988). The illusion of unique invulnerability and the use of effective contraception. *Personality & Social Psychology Bulletin*, 14, 264–270.
- Cameron, C., & Trivedi, P. (1998). *The analysis of count data*. New York: Cambridge University Press.
- Carver, C. S., & Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology*, 17, 545–568.

- Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.
- Christensen, C. M., & Bower, J. L. (1996). Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal*, *17*, 197–218.
- Clapham, M. M. (1998). A comparison of assessor and self dimension ratings in an advanced management assessment centre. *Journal of Occupational and Organizational Psychology*, *71*, 193–203.
- Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. New York: Prentice Hall.
- Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality & Social Psychology*, *57*, 1082–1090.
- Dweck, C. S., & Leggett, E. L. (1988). A social cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273.
- Elsbach, K. D., & Kramer, R. M. (1996). Members' responses to organizational identity threats: Encountering and countering the Business Week rankings. *Administrative Science Quarterly*, *41*, 442–476.
- Farh, J. L., & Dobbins, G. H. (1989). Effects of self-esteem on leniency bias in self-reports of performance—A structural equation model analysis. *Personnel Psychology*, *42*, 835–850.
- Greve, H. R. (1998). Performance, aspirations and risky organizational change. *Administrative Science Quarterly*, *44*, 58–86.
- Greve, H. R. (2003a). *Organizational learning from performance feedback: A behavioral perspective on innovation and change*. Cambridge: Cambridge University Press.
- Greve, H. R. (2003b). A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Academy of Management Journal*, *46*, 685–702.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self supervisor, self peer, and peer supervisor ratings. *Personnel Psychology*, *41*, 43–62.
- Hauser, J. R. (1998). Research, development, and engineering metrics. *Management Science*, *44*, 1670–1689.
- Hausman, J., Hall, B., & Griliches, Z. (1984). Econometric models for count data with an application to the patents—R&D relationship. *Econometrica*, *52*, 909–938.
- Huber, V. L. (1991). Comparison of supervisor-incumbent and female-male multidimensional job evaluation ratings. *Journal of Applied Psychology*, *76*, 115–121.
- Inzana, R. (2004, March 1). Disney leaders cite company's strength. *USA Today*.
- Johns, G. (1999). A multi-level theory of self-serving behavior in and by organizations. In R. I. Sutton & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 21, pp. 1–38). Greenwich, CT: JAI Press.
- Josephs, R. A., Larrick, R. P., Steele, C. M., & Nisbett, R. E. (1992). Protecting the self from the negative consequences of risky decisions. *Journal of Personality and Social Psychology*, *62*, 26–37.
- Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model comparison approach*. San Diego: Harcourt Brace Jovanovich.
- Kahneman, D. (1992). Reference points, anchors, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*, 296–312.
- Klein, H. J. (1989). An integrated control theory model of work motivation. *Academy of Management Review*, *14*, 150–172.
- Lant, T. K., Milliken, F. J., & Batra, B. (1992). The role of managerial learning and interpretation in strategic persistence and reorientation. *Strategic Management Journal*, *13*, 585–608.
- Lant, T. K., & Montgomery, D. (1987). Learning from strategic success and failure. *Journal of Business Research*, *15*, 503–518.
- Lord, R. G., & Levy, P. E. (1994). Moving from cognition to action—A control-theory perspective. *Applied Psychology—An International Review*, *43*, 335–398.
- Marinova, D. (2004). Actualizing innovation effort: The impact of knowledge diffusion in a dynamic system of competition. *Journal of Marketing*, *68*, 1–20.
- McFarlin, D. B., & Blascovich, J. (1981). Effects of self-esteem and performance feedback on future affect preferences and cognitive expectations. *Journal of Personality and Social Psychology*, *40*, 521–531.
- Meyer, M. W., & Gupta, V. (1994). The performance paradox. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 16, pp. 309–369). Greenwich, CT: JAI Press.
- Meyer, M. W. (2002). *Rethinking performance measurement*. Cambridge, UK: Cambridge University Press.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston, Inc.
- Ocasio, W. (1995). The enactment of economic adversity—a reconciliation of theories of failure-induced change and threat-rigidity. *Research in Organizational Behavior*, *17*, 287–331.
- Polzer, J. T., Kramer, R. M., & Neale, M. A. (1997). Positive illusions about oneself and one's group: Antecedents and consequences. *Small Group Research*, *28*, 243–266.
- Porac, J. F., Wade, J. B., & Pollock, T. G. (1999). Industry categories and the politics of the comparable firm in CEO compensation. *Administrative Science Quarterly*, *44*, 112–144.
- Powers, W. T. (1973). *Behavior: The control of perception*. Chicago: Aldine.
- Richards, E. (1990, July 19). Fastest growing U.S. company ever: High tech firm heads for record \$1 billion sales in first four years. *Washington Post*.
- Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. New York: Cambridge University Press.
- Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology*, *29*, 209–269.
- Short, J. C., & Palmer, T. B. (2003). Organizational performance referents: An empirical examination of their content and influences. *Organizational Behavior and Human Decision Processes*, *9*, 209–224.
- Staw, B. M., McKechnie, P. I., & Puffer, S. M. (1983). The justification of organizational performance. *Administrative Science Quarterly*, *28*, 582–600.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193–210.
- Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003). Portrait of the self-enhancer: Well adjusted and well liked or maladjusted and friendless? *Journal of Personality and Social Psychology*, *84*, 165–176.
- Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review*, *96*, 569–575.
- Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 181–227). New York: Academic Press.
- The Business Week 50: The Best Performers. (2004, April 5). *Business Week*.
- Vancouver, J. B. (2005). The depth of history and explanation as benefit and bane for psychological control theories. *Journal of Applied Psychology*, *90*, 38–52.
- Weiner, N. (1948). *Cybernetics: Control and communication in the animal and the machine*. Cambridge: MIT Press.
- Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, *90*, 245–271.
- Wood, R., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of Personality and Social Psychology*, *56*, 407–415.