

**KNOWING YOUR PLACE: SOCIAL PERFORMANCE FEEDBACK  
IN GOOD TIMES & BAD TIMES**

**Thomas P. Moliterno**

Isenberg School of Management  
University of Massachusetts Amherst  
Amherst, MA 01003  
moliterno@isenberg.umass.edu

(please direct correspondence regarding this manuscript to this author)

**Nikolaus Beck**

University of Lugano  
Faculty of Economics  
Via Giuseppe Buffi 13  
CH-6904 Lugano  
nikolaus.beck@usi.ch

**Christine M. Beckman**

University of California, Irvine  
and Robert H. Smith School of Business  
University of Maryland  
College Park, MD  
cbeckman@rhsmith.umd.edu

**Mark Meyer**

Institute for Economic Structures Research (gws)  
Heinrichstr. 30  
D-49080 Osnabrück  
m.meyer@gws-os.de

We thank Senior Editor Zur Shapira and the anonymous reviewers for their feedback. Patrick Gagliardini provided helpful methodological suggestions, and seminar participants at Aalto University, the University of Lugano, and Washington University in St. Louis shared constructive input. Earlier versions of this paper were presented at the Carnegie School of Organizational Learning Conference and the Annual Meeting of the Academy of Management.

**KEYWORDS:** performance feedback, aspirations, organizational change, behavioral theory of the firm

## **Knowing Your Place: Social Performance Feedback in Good Times & Bad Times**

### **Abstract**

Performance comparisons, specifically performance relative to aspirations, are central to the Behavioral Theory of the Firm. Firms evaluate their performance in relation to their own prior performance (“historical comparison”) and the performance of other organizations (“social comparison”) and base subsequent organizational change on this performance feedback. Of the two, social performance comparison has received relatively little theoretical or empirical development. This paper seeks to fill that gap by extending the theoretical conceptualization and empirical specification of the socially-derived performance targets against which organizations compare their performance. Drawing on insights from the social psychology literature, we argue first that organizational decision-makers monitor two socially-derived performance benchmarks: an upwardly focused “top performance threshold” marking the highest levels of performance in the reference group, and a downwardly anchored “reference group threshold” marking the performance level below which organizations can not consider themselves members of the reference group. Building on these arguments, we also motivate a new, and more complete, way to conceptualize performance comparison. Integrating socially and historically derived sources of performance feedback, we propose the “historically-based social aspiration threshold” (HiBSAT) as an additional aspiration point representing the socially-derived performance threshold closest to the organization’s prior performance. In an empirical analysis of German soccer Bundesliga clubs between 1992 and 2004, we find that organizations have both upward and downward socially-derived performance targets, and that performance relative to the HiBSAT is particularly salient in motivating organizational change.

## Introduction

A proposition central to the Behavioral Theory of the Firm (BTF) is that performance feedback is an important input to the decision process (Cyert and March 1963). Boundedly rational organizational decision-makers scan the environment, evaluate their firm's performance in relation to its prior performance ("historical performance comparison") as well as the performance of a set of other directly comparable organizations ("social performance comparison"), and base subsequent organizational adaptation on this performance feedback (Cyert and March 1963, March and Shapira 1987, 1992, Greve 2003b). This generalized model of performance feedback depends on meaningful performance benchmarks – referred to as "aspiration levels"<sup>1</sup> – against which decision-makers compare the firm's actual performance.

We focus here on social performance feedback and suggest that the nature and specification of this performance benchmark remains somewhat under-theorized (see Greve 2003b, Shinkle 2012 for reviews). The social aspiration level has generally been specified empirically as the mean performance of an industry-wide reference group (e.g., Bromiley 1991, Greve 1998, Miller and Chen 2004, Audia and Greve 2006, Baum and Dahlin 2007). While this empirical specification has yielded important insights, we join with prior researchers who have questioned this operationalization (Gooding, Goel, and Wiseman 1996, Wiseman and Bromiley 1996, Lehner 2000, Greve 2003b, Labianca, Fairbank, Andrevski, and Parzen 2009, Washburn and Bromiley 2012). Accordingly, in this paper we seek to advance the theoretical conceptualization and empirical specification of the social aspiration level by considering two distinct socially-derived performance thresholds against which organizational decision-makers benchmark their firm's performance. We then examine how these thresholds become component to an additional aspiration level that integrates social and historical performance feedback.

We begin by following researchers who have demonstrated that organizational decision-makers

---

<sup>1</sup>Here, and throughout, we follow Mezias, Chen, and Murphy (2002: 1285) in using the term "aspiration level" to refer to the number of closely related concepts (e.g., goal, reference point, performance target, etc.) that appear in the literature.

construct meaningful reference groups that are smaller than the whole industry (Fiegenbaum and Thomas 1990, 1995), and observe that these groups serve both “normative” and “comparative” functions (Kelley 1952). The normative function establishes the standards for group membership and determines what we introduce as the socially-determined “reference group threshold,” conceptualized as the performance level below which decision-makers cannot legitimately claim membership in a socially desirable group (Festinger 1954). Our theorizing in this regard is distinct but not unrelated to the stream within the performance feedback literature that has found robust support for a “survival threshold” (Boyle and Shapira 2012), and has demonstrated that proximity to this threshold affects managerial decision-making (March and Shapira 1987, 1992, Audia and Greve 2006, Chen and Miller 2007). Inasmuch as the normative function of reference groups is to establish the standards of group membership, continued participation in a particular group is conceptually analogous to “survival:” organizations falling below this threshold can no longer claim membership in the reference group, even though they may not risk death or financial failure.

The second function of reference groups, the comparative function, leads decision-makers to benchmark performance against the other firms in the group. Drawing on the social psychology literature, we propose that this comparative function leads decision-makers to aspire to perform better than others in their group and attend to top performance benchmarks within their group (Hoffman, Festinger, and Lawrence 1954, Garcia, Tor, and Gonzalez 2006). Boyle and Shapira refer to this as “the leader’s position” and argue that this is the socially-derived performance threshold which has salience for most competitors (2012: 1002). We seek to generalize the idea of a leader’s position to more complex competitive organizational contexts, where high performing organizations are grouped together (Brenner, Rottenstreich, and Sood 1999, Isaac and Schindler IN PRESS) and viewed collectively as “top performers.”

In addition to arguing for distinct social performance thresholds which manifest the normative and comparative functions of reference groups, in this paper we also align with recent scholarship that has wrestled with how sources of performance feedback combine to establish overarching organizational

performance aspirations (Washburn and Bromiley 2012, and see also Lehner 2000, Mezas et al. 2002, Short and Palmer 2003). While Cyert and March (1963) posit a general model of performance aspirations that comprises both historical and social performance feedback, subsequent research has not come to a consensus on how to make this combination (Shinkle 2012), with scholars leveraging a number of different conceptual and empirical strategies (Washburn and Bromiley 2012). Accordingly, we extend our initial arguments on refinements to social performance comparison to propose a new empirically and theoretically robust approach to integrating social and historical performance feedback into a single performance benchmark. In particular, we argue that decision-makers benchmark performance on the social performance threshold most proximal to the organization's historical performance. We refer to this as the historically-based social aspiration threshold (HiBSAT).

Thus, in this paper we seek to develop our understanding of socially-derived performance feedback first by considering social performance thresholds at the bottom and top of the organization's reference group, and then by exploring how these social performance thresholds are combined with historical performance to create an integrated benchmark for performance comparison. Following performance feedback theory (Greve 2003b) in the behavioral tradition (Cyert and March 1963), we explore these questions empirically in an analysis of organizational adaptation. In particular, we explore how social performance feedback is associated with two kinds of organizational adaptation—player line-up change and managerial change—in German Bundesliga football (soccer). There are a number of empirical benefits of this setting. First, this context allows for the empirical specification of an unambiguous competitive group. Second, German football mirrors other tournament settings and scale economies in that top performance has significant benefits. However, unlike many U.S. sport contexts, Bundesliga football also has a clear reference group threshold with the bottom performing teams in each year being “relegated” to play in a lower level/status league the following year. Thus the Bundesliga is a social reference group that serves both a normative and comparative function (Kelley 1952): participants seek both to maintain membership in the group and compare themselves to other organizations in it.

## Theory & Hypotheses

Research in the strategic management literature has long held that reference groups are smaller than the whole industry (for recent reviews see Panagiotou 2007, Schimmer and Brauer 2012). Hunt (1972) first suggested that “strategic groups” existed within the appliance industry, and the seminal contributions in this stream were made by Fiegenbaum and colleagues (Fiegenbaum and Thomas 1990, 1995, Fiegenbaum, Hart, and Schendel 1996, Shoham and Fiegenbaum 2002). They argued conceptually, and demonstrated empirically, that not only do firms cluster into strategic groups that are considerably smaller in number than the industry at large, but also that those groups serve as a reference for firm-level strategic decision making.<sup>2</sup> More recently, Panagiotou (2007) confirmed these propositions, finding that individual managers use their competitive groups as benchmarks for adjusting their firm’s competitive strategy.

Similar insights can be found in the social psychology literature. In the seminal specification of social comparison theory, Festinger (1954: 121) observed that actors place a “self-imposed restriction” on the comparison set, with the primary criterion for reference group inclusion being similarity on certain dimensions (Wood 1989: 236). One of the most relevant insights from this literature is that reference groups can serve normative and comparative functions (Kelley 1952). The former suggests that individuals seek to gain and maintain membership in a particular group, whereas the latter suggests that individuals use the group as a benchmark against which to compare themselves. Of course, much of the strategic groups literature outlined above concerns the comparative function of reference groups. However, research has not focused as much attention on the normative function of reference groups: the motivation to gain (and maintain) membership in a socially desirable group. Accordingly, our exploration of social performance comparison begins with this normative function.

### Social Performance Thresholds

---

<sup>2</sup>Subsequent work in this domain explored managerial cognition and organizational categorization (e.g., Porac, Thomas, and Baden-Fuller 1989, Reger and Huff 1993) as well as managerial perceptions of subgroups within an industry (Fombrun and Zajac 1987) as mechanisms affecting the formation of competitive groupings.

Prior research on performance feedback has explored a “failure” or “survival” point, conceptualized as a low level of performance threatening the firm’s on-going existence. March and Shapira first examined this “survival level” in their model of variable risk preferences (March and Shapira 1987, 1992), and insights regarding shifting attention to survival in the region of organizational failure have found considerable support at the organizational (Miller and Chen 2004, Audia and Greve 2006, Chen and Miller 2007) and individual (Boyle and Shapira 2012) levels of analysis.

In considering the normative function of reference groups, we propose that maintaining membership in a reference group is analogous to “failure,” although here the implication is normative—referring to an inability to participate in a particular strategic group—rather than actual organizational death. Scholars examining strategic reference groups (e.g., Fiegenbaum and Thomas 1990, Fiegenbaum et al. 1996) observe that clear performance differentials characterize reference groups. Thus, there likely exists a performance threshold below which decision-makers in organizations can no longer consider themselves members of the reference group. Moreover, to the degree that membership in the reference group is socially desirable and/or confers status benefits, falling out of the reference group might represent a significant cognitive threat for the decision-maker, given the managerial tendency toward self-enhancement (Jordan and Audia 2012). In this way, the normative function of reference groups aligns with the idea of a “competence-based” tournament, where participants are primarily focused on not losing (i.e., “the reward is punishment avoidance;” Connelly, Tihanyi, Crook, and Gangloff 2014: 40). Thus, there exists a socially-derived threshold of low performance below which the organization loses membership in the reference group. This can be an important source of negative performance feedback if organizational decision-makers have a meaningful cognitive association with the group (Porac et al. 1989, Reger and Huff 1993). We refer to the performance level below which an organization loses membership as the “reference group threshold.”

These arguments suggest two important implications. First, the reference group threshold will be contextually derived. To illustrate this proposition, we draw a familiar example from the business school rankings (Elsbach and Kramer 1996, Corley and Gioia 2000, Gioia and Corley 2002, Labianca et al.

2009) where the Top 50 business schools might comprise a distinct reference group. Accordingly, a business school can no longer consider itself a member of this group if it falls in the rankings from, say, 49<sup>th</sup> to 51<sup>st</sup>. Thus, while falling out of the reference group for the Top 50 business schools certainly does not indicate organizational financial failure, it does have significance in the normative sense: the organization is no longer a member of a desirable reference group.<sup>3</sup> Second, the normative function of reference groups implies that proximity to this threshold will be interpreted as negative performance feedback. In other words, membership in the reference group is at risk when performance hovers near the reference group threshold. Consistent with this idea, Bothner, Kang, and Stuart (2007) find that race car drivers take more risks when they are being “crowded from below,” which is the moment when their displacement from the top group is more likely.

A central tenet of BTF is that performance that fails to meet aspirations leads to problemistic search and results in organizational change and adaptation (March and Simon 1958, Cyert and March 1963, Lant 1992, Greve 2003b). In this way, we propose that the normative function of the reference group motivates organizational decision-makers to interpret a low position in the reference group as negative performance feedback and should trigger problemistic search resulting in organizational adaptation. Although scholars examining low levels of organizational performance have hypothesized a threat-rigidity effect (Staw, Sandelands, and Dutton 1981), whereby firms become paralyzed when faced with potential failure, it is important to reiterate that the “failure” event we are interested in here is normative (i.e., maintaining membership) rather than actual (i.e., organizational death). As a result, we expect low performance within the reference group to lead to problemistic search rather than threat rigidity (Ketchen and Palmer 1999). Indeed, Anderson and Brandt (1939) demonstrate that individuals in the bottom quartile of the performance distribution (whose existence was surely not threatened by the low level of performance) set much more aggressive goals for their subsequent performance than did better performing individuals. Taken together, these arguments lead us to predict a negative association between

---

<sup>3</sup>Of course, inasmuch as the reference group threshold is contextually derived, it should be validated within that context. It might well be that the Top 25 business school comprise a distinct reference group, in which case the 25<sup>th</sup> rank would represent the reference group threshold.



organizational change and the distance an organization's achieved performance is above the reference group threshold (i.e., a positive distance): an organization will make less change as its position improves relative to the threshold.

*HYPOTHESIS 1. There is a negative association between organizational change and the distance of organizational performance from the reference group threshold.*

We now focus on the comparative function of reference groups (Kelley 1952), and consider the level of performance that serves as a meaningful basis for social comparison. Again, this has generally been modeled as the industry average (e.g., Greve 1998, Miller and Chen 2004, Audia and Greve 2006, Baum and Dahlin 2007). Despite early questions about industry mean performance (Bromiley 1991, Gooding et al. 1996, Wiseman and Bromiley 1996), it is only recently that we have seen a move in the literature toward exploring alternative benchmarks for social performance comparison. For example, recent research has considered "striving" comparisons where attention is focused on those of higher status or performance (Labianca et al. 2009, DiPrete, Eirich, and Pittinsky 2010). In this way, Boyle and Shapira (2012) make considerable conceptual progress by illustrating empirically that in competitive contexts the leader's position is universally salient. We build on this work and propose that a high level of performance within the reference group is the focus of comparative processes. This proposition incorporates theoretical insights developed and tested in the psychology literature regarding upward social comparison and endows boundedly rational decision-makers with tractable heuristics that have external validity.

The theoretical literature on social comparison has long held that firms attend, and aspire, to the top levels of performance. Festinger (1954: 124-125) explicitly identifies a "unidirectional drive upward" in social comparisons, and notes that decision-makers seek to do "better and better." In particular, when actors are motivated to improve performance, "their first choice [for social comparison] is the most positive extreme" of their reference group (Wood 1989: 239). Furthermore, Heath, Huddart, and Lang (1999) demonstrate that the most influential reference point driving individual behavior is the maximum (not the median). In short, actors seek to do better than others in their reference group (Hoffman et al.

1954, Buunk and Gibbons 2007). Relatedly, scholars are growing to appreciate that an organization's position in an external ranking can be a powerful driver of organizational behavior (Sauder and Espeland 2009, Connelly et al. 2014) and change (Martins 2005). Thus the salience of rankings suggests that position within the reference group is a well-known and identity-relevant source of performance feedback. Accordingly, we suggest that striving for a high position in the competitive rankings should be an important aspect of the comparative function of reference groups.

Given the tendency for upward social comparisons, we suggest that boundedly rational organizational decision-makers employ a simple heuristic for social performance comparison: assess your organization's performance in relation to high performing competitors in the reference group. This is important inasmuch as work on strategic groups has argued and demonstrated that cognitive tractability is important for managers when making within-group comparisons (Fiegenbaum and Thomas 1995, Panagiotou 2007), and Greve (2003b) observes that decision-makers prefer to employ easily identified measures as the basis for firm-level social comparison. Indeed, striving for top performance is seen in the messages managers receive every day (Clever 2003). Leading consulting firms extol that high performers need to outperform the competition, and the business press exalts benchmarking and best practices in order to become the best in class (Hammer 1996). In short, notions of competition and "winning" by out-performing peers are almost axiomatic in Western industrial settings. Indeed, Festinger noted that "in Western culture ... there is a value set on doing better and better which means that the higher the score on performance, the more desirable it is" (1954: 124-125).<sup>4</sup> Moreover, top performance serves as a clear and unequivocal benchmark on which to assess the organization's current performance. In the rankings oriented world that we live in, organizations are constantly attending to the highest performers.

Upwardly focused social performance aspirations to obtain the highest position in the competitive

---

<sup>4</sup>Further evidence of the tendency and desire to create and monitor rankings is common and easy to find: consider rankings such as *The Fortune 500*, *The Forbes 500*, *Fortune's 100 Fastest Growing Companies*, *100 Best Companies for Working Mothers*, the J.D. Powers Awards, *Consumer Reports*, etc. In this vein, a sample of upwardly focused social aspirations in the mission statements of the *Fortune 100* is available from the authors.

rankings can be seen in the *Jeopardy!* Tournament of Champions (Boyle and Shapira 2012), but how does this generalize to organizational competitive contexts, and those with more than three competitors<sup>5</sup>? The cognitive processes that are central to current thinking on competitive reference groups (Fiegenbaum and Thomas 1995, Panagiotou 2007) as well as social psychological findings about the importance of proximity to standards or thresholds for driving competitive behavior (Garcia et al. 2006), suggest that in larger competitive contexts decision-makers are likely to see “top” performance not as the singular leader’s position, but rather as a domain of high performance within the competitive group. Indeed, both Garcia et al. (2006) and Isaac and Schindler (IN PRESS) demonstrate that individuals recognize standards that divide rankings and that the thresholds that demark these groups have particular cognitive salience. Consider again the competitive group that comprises the “Top 50” business schools. Within this competitive group, all participants recognize the importance of that list (Martins 2005). However there are multiple socially-derived performance thresholds at the top of this competitive group (e.g., the Top 10 and Top 5) that are identified by all schools (Corley and Gioia 2000) and individuals (Isaac and Schindler IN PRESS) as partitioning areas of top performance. More generally, this aligns with Washburn and Bromiley’s observation that there are “multiple peer reference points” (2012: 915). Importantly, these thresholds are the standards that serve as the basis for upwardly biased social comparison: it is the proximity to the threshold rather than the ranking itself that motivates the most competitive behavior (Garcia et al. 2006). However, there is a lower boundary to what can legitimately be considered the domain of top performance in this, and other, competitive contexts: we refer to this as the “top performance threshold.” For example, among business schools in the example above, the 10<sup>th</sup> position in the rankings may be the top performance threshold

As in the case of the reference group threshold, the exact level of the top performance threshold will be contextually derived, and subject to empirical validation within that context. Nonetheless, we suggest that all competitors strive for the top performance threshold, and that this threshold has salience

---

<sup>5</sup>Indeed, Boyle and Shapira intimate this is a direction for future research (2012: 1110). In the literature that examines tournaments, this is referred to as increasing both the “breadth” and “depth” of the tournament (Connelly et al. 2014).

with respect to decision-making inasmuch as it demarks the boundary to the domain of top performance within the competitive group. When an organization's performance falls short of this upwardly anchored top performance threshold, the comparative function of the reference group leads organizational decision-makers to interpret this shortfall as negative performance feedback. Accordingly, and following the BTF literature, we expect an organization will make more change the further its performance is below the top performance threshold, and less change the further it is above: the increasingly negative performance differential when performance is below this threshold should result in more problemistic search, just as the increasingly positive performance differential when performance is above this threshold should result in less search. Thus, we predict that there is a negative association between organizational change and the distance between achieved performance and the top performance threshold, regardless of whether performance falls above or below this threshold.<sup>6</sup>

*HYPOTHESIS 2. There is a negative association between organizational change and the distance of organizational performance from the top performance threshold.*

### **Historically-Based Social Aspiration Threshold**

Thus far we have argued for two distinct social performance thresholds: one representing the reference group threshold and one demarcating the domain of top performance. While the reference group and top performance thresholds can be regarded as socially-derived performance benchmarks associated respectively with the normative and comparative functions of the reference group, it is unlikely that both performance thresholds have the equal salience for all organizations at all times. In other words, while all organizations in the reference group might identify and be aware of these benchmarks, it is unlikely that for a given organization at a given point in time each will serve as an equally salient benchmark for performance feedback. For example, if an organization is performing close to the bottom of the reference group and is focused on the normative implications of the reference group threshold, it is unlikely that the

---

<sup>6</sup>It is important to pause and note that this prediction is not simply a restatement of Hypothesis 1. Rather, we offer two theoretically and empirically distinct reference points, and both thresholds are salient to decision-makers. Empirically, distance from the threshold for top performance does not necessarily imply proximity to the reference group threshold. While we fully explain this empirical fact in our Methods section, the crux of the matter is that ordinal performance rankings are based on an underlying continuous performance metric.

organization's decision-makers will focus equally on the performance feedback derived from the considerably greater distance from the top performance threshold.

This reasoning aligns with the renewed attention in more complex models of combining multiple sources of performance feedback into a single aspiration level (Shinkle 2012, Washburn and Bromiley 2012). Indeed, Cyert and March's (1963) formulation of the aspiration level combined the firm's prior aspiration level, the firm's prior performance, and the performance of comparable organizations, and subsequent scholarship has wrestled with the best way to make this combination (Bromiley 1991, Mezas et al. 2002, Greve 2003a, Baum, Rowley, Shipilov, and Chuang 2005). Recently, Hu, Blettner, and Bettis (2011) offer evidence from simulation research to support a three-reference-point preference function. In their rigorous treatment of how historical and social performance feedback are combined, Washburn and Bromiley (2012) find support for a "varying parameters model," demonstrating that performance improvement relative to one benchmark may have different salience to decision-makers as a function of how their organization is performing relative to other benchmarks.

However, the varying parameters model proposed by Washburn and Bromiley also raises a number of provocative questions. They note that the "possibility that weights vary depending on performance relative to a reference point raises the question: what reference point?" (2012: 908). Their answer proposes that social comparison determines the initial reference point and they use the industry median as their starting point. While this specification is consistent with the initial model proposed in Bromiley (1991), it is not based—as the authors note—on strong theoretical argument. Indeed, Short and Palmer (2003) also propose and find that decision-makers use multiple performance benchmarks, but conclude that internal (i.e., historical performance) benchmarks are more salient than external (i.e., social performance) comparisons. We approach this question by building on the theoretical basis for two social performance thresholds, and propose that managers focus their attention (Ocasio 1997) on the social threshold most proximal to the firm's historic performance. This aligns with Washburn and Bromiley's observation that the aspiration level should be perceived as "achievable" and is influenced by the firm's current performance (2012: 908). Thus, the organizational decision-maker's relative focus on either the

top performance or reference group threshold should depend, in part, on the firm's actual past performance. We expect, then, that decision-makers integrate their organization's historical performance with the more proximal social performance threshold and consider most salient the social performance threshold more proximate to the firm's historical performance. We label this the "historically-based social aspiration threshold" (HiBSAT).

To illustrate the HiBSAT, we return a final time to the business school example. Our proposed HiBSAT suggests that the social performance threshold that is more salient to a given school will depend in part on the prior ranking of the school. Thus, the dean of a business school historically ranked in the mid-forties is more likely to aspire to maintain membership in the Top 50 than to cross into the domain of top performance by achieving a Top 10 ranking (although she no doubt identifies the Top 10 as an important top performance threshold). Conversely, a business school historically ranked in the mid-teens is more likely to aspire to the top performance threshold of the Top 10. Importantly, while not considering aspirations or performance thresholds *per se*, Martins (2005) demonstrates evidence that corroborates the comparative logic underlying the HiBSAT—specifically in the context of business school rankings—finding that schools compare their current ranking with existing perceptions of the school's status.

In this way, the HiBSAT proposes that organizational decision-makers combine historical and social performance feedback into a single performance benchmark: the organization's historical performance focuses attention on a particular social performance threshold. Stated differently, decision-makers know their organization's historical place in the performance distribution—whether good or bad—and focus attention on the social performance threshold most proximate to that position as the more salient benchmark for performance feedback. Consistent, then, with the BTF's problemistic search mechanism linking performance feedback and organizational change and adaptation, we predict that there is a negative association between organizational change and the distance between achieved performance and the HiBSAT. Organizations will make more change as performance falls from the HiBSAT indicating a problem that needs to be solved, and will make less change as performance increases relative to the HiBSAT point, indicating better than expected performance. Following research in this tradition (Greve

2003b), we formalize this prediction as two separate effects. This allows us to empirically examine whether the strength of the proposed relationship differs for those above and below the HiBSAT.

HYPOTHESIS 3a: *An organization will make more change the further it is below the HiBSAT.*

HYPOTHESIS 3b: *An organization will make less change the further it is above the HiBSAT.*

## Methods

### Data & Context

We test these hypotheses using a comprehensive dataset of German Bundesliga football (soccer) teams from August 1992 to May 2003. The Bundesliga is the most profitable sector of the German sports industry, with 2004/2005 average team revenue exceeding €71 million (Deutsche Fußball Liga GmbH, 2006). Importantly, the Bundesliga offers an ideal context for studying performance aspirations in a competitive market since the boundary of the reference group is clear: every Bundesliga team knows and monitors the other teams against which it competes in the Bundesliga. Moreover, performance can be determined in a fully objective manner, thereby facilitating examination of the comparative function of the reference group. Finally, the process of “relegation,” whereby the three worst performing teams in the Bundesliga at the end of the season are replaced for the next season by the three best performing teams in the lower-level “second Bundesliga”<sup>7</sup>, provides an ideally suited empirical context in which to examine the normative function of the reference group. While the three teams that are relegated to exit the Bundesliga “survive” in that they continue to function as viable enterprises playing in the lower-level league, they clearly have lost their membership in the reference group.

During our observation period, 18 teams competed annually in the Bundesliga, during a season that ran from late summer to spring.<sup>8</sup> The Bundesliga season is divided into 34 multiday *Spieltage*—or “rounds”—during which all teams play (i.e., the 18 teams play a total of 9 matches in each round). Consequently, a total of 306 matches are played per season, providing data on a total of 3,366 Bundesliga

---

<sup>7</sup> These were the rules of relegation in the Bundesliga during the years in our sample: these rules changed in other years. Relegation occurs similarly in other professional football leagues.

<sup>8</sup> Here, and throughout this section, our discussion of Bundesliga and UEFA rules and regulations, and their impact on the variable construction, is dramatically constrained in the interest of space. A full and comprehensive treatment of the regulations that are relevant for our analysis can be provided by the authors.

matches (i.e., “games”) over the 11 seasons we studied. Since our dataset measures each match twice, once for the team that plays at home and once for the team that plays away, this results in 6,732 team-level observations.<sup>9</sup> Due to the end-of-season relegation system, a total of 32 different teams joined the Bundesliga during the 11 year period. A team that relegated to a lower league and subsequently returned to the Bundesliga was not included in our dataset during the years in which it competed in a lower league.

As we discuss below, our dependent variables measure two different team-level changes, namely lineup and coach change, prior to the start of each round during a season. Since the period between two consecutive seasons is often marked by a considerable reconfiguration of a team’s personnel due to expiring contracts and the recruitment of new players or coaches, we dropped all first round observations from our dataset. As a consequence, our final data set consisted of 6,334 team-level observations. For each match, a complete record of various statistics on the whole team and individual team members was acquired from Sport-Dienst-Agentur Merk, a commercial service agency which compiles a wide variety of football data.

### **Dependent Variables**

We are theoretically interested in organizational adaptation and change. Empirically, we operationalized change in two ways: *lineup change* and *coach change*. These two dependent variables capture operational changes to the configuration of the organization’s human capital (e.g., players) and replacement of the manager tasked with making these human capital configurations (e.g., coaches).

The dependent variable *lineup change* is a count variable of the changes between a team’s current initial lineup of players and the players that played in the previous match. That is, this dependent variable is the number of new players a given team uses in the current match. Higher levels of this variable necessarily suggest the intentional decision to undertake more change and adaptation to the strategically important human capital resource (Ployhart and Moliterno 2011; Nyberg, Moliterno, Hale, & Lepak 2014) deployed by the team.

---

<sup>9</sup> Bad weather can cause matches in the Bundesliga to be postponed until later in the season. We took postponement of matches into account and ordered all matches according to the actual date when they were played.



It is important to note that this variable is not simply a comparison of a team's *initial* lineups for two consecutive matches. In the Bundesliga up to three field players and the goalkeeper can be substituted during a match. As a result, a player who entered the game as a substitute in the previous match and appears as a member of the *initial* lineup in the next game does not arguably represent lineup change that might be determined by performance feedback as we conceptualize it here. We therefore counted only the number of new players in the initial lineup who did not enter the match as a substitute in the previous match. This necessarily also accounts for lineup changes that occur during a match because a player is injured and another player comes in to replace him. In addition, we exclude from the calculation of *lineup change* any player who enters the lineup due to another player being removed from the prior match due to receiving a "red card" or to having accumulated several yellow cards over several matches.<sup>10</sup> Our results are robust to an alternative model specification where we included the count of red cards that affected the team's line-up as a control variable rather than subtracting it from the measure of lineup change.

*Coach change* captures change that is focused more directly on the managerial resources (Adner and Helfat 2003) that are tasked with the team's operational human capital decisions (i.e., the field manager charged with creating the lineup and making in-game player assignment decisions). To measure this kind of organizational adaptation and change, we coded the dummy variable *coach change*, to capture when a team replaced its coach. When the coach differed from the coach in the previous match we coded *coach change* with 1, otherwise with 0. We observed 90 coach changes during our sampling frame.

### Independent Variables

In the Bundesliga, as in all European football leagues, a team's performance at any point in time is measured in number of points accumulated up to that point in the season. Prior to the 1994/5 season,

---

<sup>10</sup> In football, "red" and "yellow cards" are given to players who have committed a serious foul or behaved in an unsportsmanlike manner: the red card is the more severe punishment. When a player receives a red card, he is taken out of the game and his team is not allowed to substitute in a new player to replace him for that game (i.e., the team plays shorthanded). Moreover, in the Bundesliga, players who receive a red card are banned (i.e., they have to stay out of the lineup) for at least one additional match, although in this subsequent match the team is not required to play shorthanded (i.e., they substitute for the banned player). Players are also banned for one match when they have accumulated a certain number of yellow cards during a season (the number of accumulated yellow cards varied over our observation period). Given these rules, teams with players that have received red and yellow cards necessarily require lineup changes in the next match and we therefore excluded those changes from the dependent variable.

two points were awarded to the winner of a match and one point to both teams if a match resulted in a draw. From 1995/96 onward, the points awarded to the winner of a match increased to three. Like win percentage in American sports (Olson and Schwab 2000, Berman, Down, and Hill 2002, Moliterno and Wiersema 2007, Sirmon, Gove, and Hitt 2008, Holcomb, Holmes, and Connelly 2009), point totals are an intuitive measure of operational performance in professional football, and in a study of performance aspirations, it is important to measure context-specific variables of relevance to decision-makers (Greve 2003b). This is a benefit of the Bundesliga setting, since in other industrial contexts the firm's decision-makers might focus on any of a number of operational or financial performance metrics (e.g., market share, profitability, ROA, etc.), making it difficult to know the appropriate performance measure for social comparisons.

Point totals determine teams' ranking in the overall standings: this provides another empirical advantage. Hypotheses 1 and 2 consider, respectively, the effect on change of performance feedback when an organization's performance draws near the reference group threshold or recedes from the top performance threshold. Taken together, Hypotheses 1 and 2 specifically suggest that it is possible for an organization to recede from the threshold for top performance without simultaneously drawing closer to the reference group threshold. This is owing to the fact that ordinal rankings are based on an underlying continuous performance metric. In the Bundesliga, this underlying performance metric is the point total. The threshold for top performance and the reference group threshold are expressed by certain *ranks* in the point total standings (described below) which makes it possible to measure the distance from these thresholds in terms of *point* differences during a season. For example, for a team ranked 10<sup>th</sup>, the distance *in points* between it and the team ranked 4<sup>th</sup> might increase in any given round, even if the difference in points between the team ranked 10<sup>th</sup> and the one ranked 15<sup>th</sup> remains the same (e.g., the 4<sup>th</sup> ranked team wins its match and the 10<sup>th</sup> and 15<sup>th</sup> ranked teams lose their matches; c.f. Footnote 6).

By basing the distance measures on differences in points, it is also possible to include distance to both thresholds in a single model and thereby juxtapose the effects of both performance comparisons. This is due to the fact that even though the distance measures are correlated with each other, they are not

confounded. This ability to simultaneously consider distance from both social performance thresholds is an empirical advantage, allowing us to simultaneously test our hypothesized effects on organizational change of distance from the reference group threshold and the top performance threshold (see Boyle and Shapira 2012 for a similar approach).

*Reference Group Threshold.* In the Bundesliga, there is a clear reference group threshold: relegation to the Second Bundesliga. During our observation period, the teams that relegated were always the last three in the overall standings (i.e., teams whose point totals placed them on the 16<sup>th</sup>, 17<sup>th</sup>, and 18<sup>th</sup> ranks at the end of a season). We therefore defined the reference group threshold as the 15<sup>th</sup> rank in the overall standings of the German Bundesliga because this is the last rank that allows a team to remain in the Bundesliga.

*Top Performance Threshold.* Hypothesis 2 argues for the salience of a domain of top performance for social performance comparison. Bundesliga football teams compete not only for the league championship (i.e., the team with the greatest number of points at the end of the season), but also the ability to play in the European Champions League and/or the UEFA Cup (now called the EURO League). These European leagues are both widely followed and highly publicized, inasmuch as the competing teams were all highly ranked in their domestic leagues. Importantly, participating in these leagues is not only a source of pride for the team: it is also a source of considerable financial and status benefits. Of the two international leagues, it is easier to qualify for the UEFA Cup. While the financial and status rewards associated with participation in the UEFA Cup are somewhat less than those that attend the European Champions League participation, the benefits are still non-trivial. Without qualifying for the UEFA cup, there are no opportunities for teams to obtain the financial and status rewards of international competition. Thus the rank needed to participate in the UEFA Cup is the contextually relevant operationalization of the top performance threshold, and as such was used to create the variable operationalized to test Hypothesis 2.

To create the variables used in our hypothesis tests we calculated the distance between the performance of a focal team and the reference group threshold (Hypothesis 1) and the top performance

threshold (Hypothesis 2), after each round in the season. Again, it is important to note that we calculated the *difference in points* between the focal team and the team that occupied the threshold rank after each round, using the equation  $P_i - P_j$  (Eqn. 1), where  $P$  is a team's point total,  $i$  indexes the focal team, and  $j$  indexes the team at the performance threshold. For example, a team's performance above the top performance threshold will be an increasingly positive number whereas the distance below that threshold will be increasingly negative. Using Eqn. 1 to measure a team's performance relative to the 15<sup>th</sup> rank and the UEFA Cup threshold, we constructed the continuous variables *distance from reference group threshold* and *distance from top performance threshold*.

*Historically-Based Social Aspiration Threshold (HiBSAT)*. Hypothesis 3 predicts organizational change when an organization's performance recedes from its own HiBSAT. We constructed the distance from the HiBSAT in three steps (Table 1 illustrates these steps in the creation of this measure for one of the teams in our sample, VfB Stuttgart).

First, we follow the standard practice in the performance feedback literature and constructed the team's historic aspiration level (HAL) based on the final rank of a team at the end of a season using the exponential weighted moving average specified in Levinthal and March (1981):  $HAL_t = \alpha FinalRank_{t-1} + (1-\alpha)HAL_{t-1}$  (Eqn. 2), with  $t$  indicating seasons. Here we used *final rank*, and not *points*, as the performance measure. As an end-of-season performance outcome that becomes the basis for interseason aspiration setting, rank is more salient than points, which are better suited for intraseason tracking of performance (i.e., at the end of the season what matters is the rank the team achieves, whereas during the season what matters is how close a team is, in points, to achieving a particular rank). Following the technique reported by Baum et al. (2005), we constructed this variable with values of  $\alpha$  set at 0.25, 0.5, and 0.75, and opted for  $\alpha = 0.75$ , which had the best overall fit. Indeed, the specification of a large  $\alpha$  has face validity in this context, inasmuch as the season-specific returns to high or low performance in sports justify a "short" memory.<sup>11</sup>

---

<sup>11</sup> For teams that were relegated, and then subsequently returned to the Bundesliga, we also computed the HAL during their period of relegation. We did this in order to estimate the right aspiration threshold for them when

Second, we calculated the absolute difference between the team's HAL rank and the reference group and top performance thresholds. The threshold with the smallest absolute difference to the HAL rank was then considered to be the *historically-based social aspiration threshold (HiBSAT)* of a team for a given season. Thus teams could have "HiBSAT top performance" or "HiBSAT avoid relegation," depending on whether their HAL was closer to the top performance or reference group threshold, respectively. For example, if a team aspires, based on its own historical performance to the top performance threshold at the end of, say, the 1994-1995 season, we used HiBSAT top performance for that team for the 1995-1996 season. Maintaining HiBSAT as the performance benchmark for the whole season is both empirically and contextually valid. Goals are usually set in the interseason break and remain constant during the following season. Coaches or important players frequently dismiss expectations, alluded to by media representatives, that a team should now aspire for higher goals since it performed unexpectedly well in recent matches.<sup>12</sup> Even if the team may be "flirting" with a higher aspiration or fearing that aspirations have to be lowered during a season (both of which would normally not be communicated to the media), it is the relative standing at the end of the season that sets the baseline for the subsequent season's aspirations.

Third, for each round of a season we calculated the difference between a focal team's point total and the points of the team which currently occupies its respective HiBSAT rank. We then constructed a spline function (Greene 2008) of this variable for performance below and above the HiBSAT. In this way we calculated the variables *performance below (above) HiBSAT*, which measure the distance in points to the HiBSAT for all Bundesliga teams at each round during a given season.

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

### Control Variables

---

they returned to the Bundesliga. For these teams we set *Final Rank<sub>t-1</sub>* equal to the final rank that the team occupied in the lower league plus the 18<sup>th</sup> rank of the Bundesliga.

<sup>12</sup> An example of this kind of attitude is represented by the following news item: "'Dreaming of the UEFA Cup is not forbidden,' said goalkeeper Robert Enke after the 3-1 against Dortmund. Yet for the coach, who took over the team at the bottom of the standings, this is all a bit too fast. 'Fans have of course the right to dream. But we are realists'" (Stuttgarter Nachrichten, 26<sup>th</sup> of February 2007: 27, no author given).

We included a number of variables in our models to control for organizational changes that may be driven by contextual and/or empirical effects. To ensure that our models were a conservative test of our theory, we included *HiBSAT avoid relegation* as a dummy variable for all firms with that HiBSAT in the models. This dummy variable controls for the cohort that comprises the teams with the same low HiBSAT. Since we lack financial data, this cohort dummy controls for the lack of wealth: any team that has *HiBSAT avoid relegation* likely has fewer financial resources than better performing teams (i.e., because HiBSAT is derived based on prior performance and historically higher performing teams should have more financial resources). This lack of resources constrains the team's ability to maintain a large pool of players, allowing fewer lineup changes to experiment with new team constellations connected to alternative tactical strategies.

To capture the effect of time during the playing season, we control for the current round within a season. It is possible that over the course of the season there is a systematic trend toward greater, or fewer, changes in the player lineup. Moreover, it's entirely possible that any effect of time is non-linear given that the payoffs to changes are greater earlier in the season when teams are working out a cohesive and effective starting lineup. In addition, the return to changes in the lineup and coaches might well decrease over the course of the season. This is owing to the fact that as the season progresses, the point differential between a team and the team on its HiBSAT may become increasingly difficult to surmount. Thus towards the end of the season changes in lineup or replacing the coach arguably have less impact since it is increasingly less possible for a team to impact its final ranking, given the number of matches left to be played (and thus the total number of points possible to gain). We included a squared term of the round number variable in order to account this non-linear influence of time.

Since a team that achieves the top performance threshold during the *prior* season plays in the Champions League or UEFA Cup in the *current* year, we constructed a dummy variable that captured the participation of a team in one of these competitions during the current season. This is a contextually important control inasmuch as the need to field a team—and probably the best possible team—for a European competition would necessarily impact lineup decisions for the current season's scheduled

games. Moreover, engagement in European competitions might also have an effect on *coach change*: current participation in a European competition (earned due to top performance in the prior season) might outweigh possible current underperformance in the current season, thereby reducing the coach's risk of being terminated. Therefore, all observations of a team in a given season that fell into the period of participation in a European Cup competition were coded with values of 1 and all other observations were coded with 0.

We also accounted for the strength of a focal team's opponent in order to control for the possibility that coaches make additional lineup changes when facing an especially strong (or weak) opponent. We first calculated the difference in ranks between the focal team and the opponent. Since high ranks indicate low performance, positive values of this variable mean that the opponent is currently stronger than the focal team while negative values indicate that the focal team is currently stronger than the opponent. We also included this variable in the models of coach change.

Previous research has found that diversity in tenure, age, ethnicity and experiences hampers the development of team cohesion, thereby increasing the likelihood of change (Pfeffer 1985, Williams and O'Reilly 1998, Alesina and La Ferrara 2002), and previous work in this context has found that more heterogeneous teams have a significantly lower performance (using the same index described here; Beck and Meyer 2012). Thus, for the analysis of lineup change we also constructed a heterogeneity index for each team that consisted of six demographic characteristics of the team's players: current and overall tenure in the team, age, nationality, number of Bundesliga matches and number of titles won by each player.. Following Gibson and Vermeulen (2003) we compared each of these individual players' characteristics with every active teammate in a particular match by calculating the ratio between the lower and the higher value of the comparison. This means that we calculated for each pair of teammates six comparisons (i.e., one for each demographic characteristic). For nationality, we created a dummy variable that took the value of one for pairs of teammates with the same nationality. Then we summed up these six comparisons for each pair of teammates, which resulted in a team vector of aggregated comparisons. We calculated the mean of this vector and took the inverse of this mean as the index of heterogeneity. Finally,

we calculated a weighted moving average of this heterogeneity index. We used the same mathematical approach to construct this moving average as we used for the construction of HAL (Eqn. 2), replacing Final Rank with the heterogeneity index. We did not include this variable in the coach change models since this kind of change is carried out by the club's directors, and as such is not influenced by the composition of the team.

In the models regressing the *lineup change* dependent variable, we also controlled for the replacement of a coach. Since a new coach might have very different views on tactics, it is entirely possible that he may make a greater number of lineup changes in his first match. We therefore coded a dummy variable for the first match in which a new coach was responsible for a team. In constructing this variable, we only took within-season coach changes into account. Finally, we included the *duration of a new coach's tenure* measured in rounds until the end of a given season or the arrival of yet another coach in the current season in both the *lineup change* and the *coach change* models. In the *lineup change* models, we expect a reduction of lineup change over tenure of a coach during a season. In the *coach change* models, the inclusion of this variable controls for the possibility that coaches might gain (or lose) trust of the team's higher-level management and ownership.

### **Analysis and Estimation**

Since the number of new players in the initial lineup is a non-negative discrete variable (i.e., a count-variable), we modeled GEE Poisson regressions in our models of *Lineup Change*. Since a graphical representation of lineup change made it evident that the residuals were autocorrelated, we specified our models using a one period autoregressive (AR1) correlation structure, thereby controlling for the autocorrelation of the errors in such a way that errors of observations farther apart in time can be considered more weakly correlated than errors of observations closer to the current observation. We included team fixed effects to control for unobserved heterogeneity (Beck, Brüderl, and Woywode 2008, Greene 2008). Since we are interested in changes that are made as a function of the idiosyncratic performance of the teams, we did not want these effects to be biased by spurious influences stemming from different time-constant but team-inherent tendencies to undertake changes. Finally, in all models we



controlled for time effects by including dummy variables for each season with the 1992/1993 season as the excluded category.

In our models of *Coach Change*, we used a panel-logit model with fixed effects. We calculated robust standard errors using a Huber-White sandwich estimator. Since fixed effect logit models require that at least one change per observation unit occurs, those teams that never experienced a coach change were dropped from the analysis, resulting in 5874 observations in the *Coach Change* analysis.

## Results

Table 2 reports the correlations and descriptive statistics of all variables included in our analysis. The correlations between our distance measures range from 0.32 to 0.77, suggesting strong correlations within the typical range of studies of multiple sources of performance feedback. Table 3 reports the results of our GEE regressions of initial lineups changes with fixed-effects. Model 1 is the control only model. The first coefficient denotes the effect of the HiBSAT dummy. Consistent with the idea that those with higher HiBSAT may have accumulated more financial resources that can be used for acquiring players, we find that teams in the low *HiBSAT* reference group, i.e. those teams with aspiration to avoid relegation, make less change than those with *HiBSAT Top Performance* (the excluded category).

-----  
Insert Tables 2 & 3 about here  
-----

Results for other control variables are generally consistent with expectations. The effects of the linear and squared term of the current round indicate that during the season the amount of change first increases and then decreases. This increase in changes indicates that the coach makes more change as the season progresses. However, change then decreases likely due to the lower chances of making an important difference towards the end of the season. The positive significant coefficient for *moving average heterogeneity* means that teams with players who possess a high degree of demographical diversity make more lineup changes. We also find evidence that participation in other cup competitions results in more lineup change (due to the need to rest players and allocate them across different

competitions). A coach succession within a season leads to an increase in lineup change: not only do new coaches often have a different tactical approach than their predecessors; there is often the expectation that the newcomer is expected to make changes in how the team is managed. Finally, the difference in ranks has a significantly positive effect on lineup change, indicating that teams undertake more changes when they face opponents that are currently ranked above them in the overall standings.

Models 2-5 in Table 3 report tests our hypotheses with respect to *lineup change*. It is important to recall that our independent variables measure distance from a particular performance threshold. The interpretation of the regression coefficients is fairly straightforward when considering distance from the reference group threshold (Hypothesis 1), since small values of *distance from reference group threshold* indicates performance more proximal to an undesirable threshold. Thus, our predicted association would anticipate a negative regression coefficient for this variable. Calculation of *distance from top performance* results in increasing negative values the further a team is below this threshold and increasingly positive values the further a team is above this threshold. Thus, our predicted association here also anticipates a negative coefficient: larger negative values below the threshold should be associated with more changes, while larger positive values above the threshold should be associated with fewer changes. The same interpretation can be applied to coefficients measuring proximity to HiBSAT, since achieved performance can be either above or below this aspiration level.

Model 2, then, displays a significant negative effect of *distance from reference group threshold* ( $\beta = -0.0123$ ;  $p < 0.01$ ) indicating that as a team's performance, in points, climbs above the point total of the team that occupies this threshold (i.e., the 15<sup>th</sup> rank in the standings), the team makes fewer lineup changes. This offers preliminary support of Hypothesis 1. Model 3 reports the test of Hypothesis 2. The coefficient for *distance from top performance threshold* is also negative and significant ( $\beta = -0.0133$ ;  $p < 0.01$ ), indicating that as a team's performance, in points, falls further below (above) the point total of the team that occupies the rank demarking the threshold for top performance (the rank needed to play in the UEFA Cup), the team makes more (fewer) lineup changes. In Model 4 we enter both *distance from*

*reference group threshold* and *distance from top performance threshold*. The coefficient for *distance from top performance threshold* remains significantly negative ( $\beta = -0.0127$ ;  $p < 0.05$ ) while the effect for *distance from the reference group threshold* becomes non-significant. Thus, with respect to *lineup change* we find support for Hypothesis 2 but not for Hypothesis 1: the amount of organizational change increases the further performance falls below an upwardly bounded performance threshold demarking top performance, and decreases the further it is above it.<sup>13</sup>

Model 5 reports the test of Hypotheses 3a and 3b. The coefficients for the spline variables measuring performance above and below HiBSAT both are both negative and significant (although not significantly different from one another). Thus, Hypotheses 3a and 3b are supported: the amount of organizational change increases the further performance falls below the HiBSAT, and decreases the further it is above the HiBSAT. We conducted an additional *post hoc* analysis in an effort to unpack how HiBSAT affects organizational change. In Model 6 we include interactions between the HiBSAT dummy variables (*HiBSAT top performance*, and *HiBSAT avoid relegation*) and both measures of social performance feedback (*distance from reference group threshold* and *distance from top performance threshold*). Our theory suggests that the distance to the threshold most proximate to the HiBSAT should be most salient to the firm. This expectation holds true. For the teams in the *HiBSAT top performance* cohort, increasing distance from the top performance threshold resulted in significantly more lineup change ( $\beta = -0.0169$ ;  $p < 0.05$ ), whereas distance the reference group threshold (relegation) did not result in a significant effect. Likewise, for the group of teams in the *HiBSAT avoid relegation* cohort, proximity to that reference group threshold resulted in more lineup change ( $\beta = -0.0160$ ;  $p < 0.05$ ), whereas distance to the top performance threshold (participation in the UEFA cup) did not result in significant increase in lineup change. In total, the results of Model 6 show strong support for the theoretical proposition motivating Hypotheses 3a and 3b: organizations combine historically- and socially-derived sources of performance feedback and attend to performance feedback based on the socially-derived performance

---

<sup>13</sup> In supplemental analysis (available from the authors) we regressed *lineup change* on a traditional measure of social aspirations, operationalized as the distance from median performance. We find no support for this median specification once we account for our new measures of social aspirations.

threshold to which they have historically been closest.

-----  
Insert Table 4 about here  
-----

In Table 4 we report tests of our hypotheses using the *coach change* dependent variable. The results for regressions of *coach change* mirror to a large extent those for *lineup change*. Model 1 is the control only model with largely consistent results. The significant coefficients for *round* ( $\beta = 0.1921$ ;  $p < 0.01$ ) and *round squared* ( $\beta = -0.0043$ ;  $p < 0.01$ ) indicate a greater probability for a coach to be terminated early in the season, while later in the season this tendency decreases. Coach replacements towards the end of the season will have increasingly limited benefits. The significant positive coefficient for *difference in ranks* ( $\beta = 0.1013$ ;  $p < 0.01$ ) suggests that the probability of coach change increases with the strength of a team that a focal team has to face in the current round: matches against strong teams are seen as opportunities that facilitate the replacement of a coach.

Models 2-5 in Table 4 report tests our hypotheses with respect to *coach change*. In Model 2, *distance from the reference group threshold* has a significant, negative coefficient ( $\beta = -0.1952$ ;  $p < 0.01$ ) which indicates that coach changes become increasingly likely when a team approaches the reference group threshold. Similarly, in Model 3 the regression of *coach change* on *distance from top performance threshold* results in a significant negative association ( $\beta = -0.1614$ ;  $p < 0.001$ ), consistent with the effects reported in the Table 2 tests of Hypothesis 2.

When entering *distance from top performance threshold* and *distance from reference group threshold* in the same regression (Model 4) we observe a significant association for only *distance from the reference group threshold* ( $\beta = -0.1907$ ;  $p < 0.01$ ). Thus for coach succession, proximity to the reference group threshold is, on average, the decisive source of social performance feedback, while distance from the domain of top performance is, on average, unimportant for coach change. In interesting and striking contrast to the results reported for the *lineup change* regressions (where we found strong support for Hypothesis 2 but not Hypothesis 1), these results lead to strong support for Hypothesis 1 but not for

Hypothesis 2. A possible explanation of this finding is that the behavioral response to threat of falling out of the reference group may be to initiate a different kind of change. We reflect on this finding in the Discussion.

Finally, testing Hypotheses 3a and 3b for the *coach change* dependent variable, Model 5 demonstrates that the distance to HiBSAT is negatively and significantly associated with change both above ( $\beta = -0.3141$ ;  $p < 0.05$ ) and below ( $\beta = -0.1754$ ;  $p < 0.01$ ) HiBSAT. Thus, and consistent with our hypotheses tests using the *lineup change* dependent variable, we find support for our predictions that the amount of organizational change increases the further performance falls below the HiBSAT, and decreases the further it is above the HiBSAT. We again conducted an additional, more detailed analysis of distance to HiBSAT which is presented in Model 6. Interestingly, this model reveals that only for teams that have the *HiBSAT avoid relegation* does proximity to the reference group threshold lead to a significant increase in the likelihood to change the coach. For teams with *HiBSAT top performance* the distance from the top performance threshold has a negative effect, as expected, but it is insignificant. This may be because of the low number of *coach change* observations.

## Discussion

Performance feedback theory is concerned with how performance comparisons drive organizational decisions to engage in organizational adaptation and change, with a central argument that organizations make meaningful performance comparisons by benchmarking a reference group of comparable organizations. We both follow earlier critiques of industry mean as the measure of the social aspiration level (Gooding et al. 1996, Lehner 2000, Labianca et al. 2009, Washburn and Bromiley 2012), and draw on the long-standing proposition in the strategic management literature that firms create reference groups that are smaller than the whole industry (Fiegenbaum and Thomas 1990, 1995, Fiegenbaum et al. 1996, Shoham and Fiegenbaum 2002, Panagiotou 2007, Schimmer and Brauer 2012). To these two starting points, we add insights from cognitive and social psychology concerning how social reference groups affect behavior, and the results we report here point to an expanded theoretical

perspective on how social performance comparison is integrated into organizational change decisions.

We found empirical support for the idea that the normative function of reference groups (Kelley 1952) results in a threshold of low performance—below which organizations cannot consider themselves members of a reference group—that has salience to organizational decision-makers, and that performance declines toward this reference group threshold trigger a problemistic search mechanism and result in organizational change. However, we also found support for an important threshold at the other end of the social performance distribution. An upwardly anchored top performance threshold reflects the organizational reality that status and financial rewards generally accompany the highest levels of performance in a competitive group, and integrates findings from individual-level research (Festinger 1954, Hoffman et al. 1954, Wood 1989, Förster, Grant, Idson, and Higgins 2001, Garcia et al. 2006) into behavioral theory's perspective on how decision-makers integrate performance feedback. Thus the results we report here extend the core theoretical arguments of the performance feedback literature by integrating insights from both the strategic groups and social psychology literatures and suggesting additional social aspiration points that are relevant to organizational decision-makers. In this way, we move beyond the under-theorized suggestion of industry mean as the relevant social performance benchmark.

In integrating the idea of the normative and comparative functions of reference groups, we have extended our theoretical understanding of social performance benchmarks. For example, among Olympic competitors placing third and thereby earning a place on the medal podium, third is considered “top” performance, even if there are differential benefits to placing second or first (Medvec, Madey, and Gilovich 1995). Extending this example, consider the athletes who are finalists in a given Olympic event (e.g., the eight runners in the finals of the 100 meter dash). These athletes comprise a clear social reference group. Athletes that do not make it through the qualifying rounds to participate in the finals cannot claim membership in this reference group, even though they certainly remain elite Olympic athletes. Thus, just as there is a domain of top performance in this social reference group (i.e., the medalists in the 100 meter dash), there is also a reference group threshold marking the lower bounds of performance for inclusion in the group (i.e., the cut-off for participating in the finals).

We also offer strong empirical support for a new theoretical construct: a historically-based social aspiration threshold (HiBSAT). We argued that the relative salience of the social performance thresholds at the top and bottom of the reference group performance distribution should depend on the organization's past performance. This perspective builds upon recent scholarly work in the performance feedback literature (Washburn and Bromiley 2012) by supporting the proposition that social and historical information are combined into a single aspirational performance benchmark. However, and in an extension of prior work, we demonstrated that this benchmark varies as a function of managerial attention to socially desirable performance thresholds: decision-makers select as the aspiration point the socially-derived performance threshold that is closest to the firm's historical performance, whether that is at the top or the bottom of the performance distribution. In this way, the findings reported in this paper suggest a new way to conceptualize the aspiration level firms use to derive performance feedback. Simply put, decision-makers know their place in the competitive group, and derive performance feedback based on that place.

An interesting, and somewhat surprising, result from our analysis concerns the different pattern of effects we observe for our two different operationalizations of organizational change. Specifically, when we included distance from the reference group threshold and distance from the top performance threshold in the same model, we observed that lineup change was negatively associated with only the latter, while coach change was significantly and negatively associated only with the former. Thus, it appears that in the case of coach change, distance from the reference group threshold dominates the effect of distance from the top performance threshold, whereas the opposite is true for the case of lineup change. There may be a contextual explanation for this pattern of effects: distance from the top performance threshold might be less important, on average, with respect to coach change since such a change is often very costly. In particular, coaches that are fired before their contract ends—which is always the case for within-season dismissals—still receive their salary until their contract expires. Accordingly, an organization may be willing to undergo this considerable expense only when faced with the normative threat of falling out of the reference group. If this is contextually the case, we might then more generally argue that the

normative threat of falling from the reference group presents a greater problem and thus requires a larger and more dramatic organizational change response. Alternatively, we might consider this as a qualitatively, and not quantitatively, different response to social performance feedback. That is, making changes in how the team's player resources are configured is a fundamentally different type of change than that which is focused on the management regime that manages those resources. In this vein, Moliterno and Beckman (2011) find that the lowest performing organizations shift from changes focused on resource deployment (Sirmon et al. 2008, Holcomb et al. 2009, Sirmon and Hitt 2009) to those focused on resource management (Sirmon, Hitt, and Ireland 2007). It is possible that we find a similar effect here. While our data do not allow us to explore these intriguing issues further, it is clear that additional research is warranted to help us better understand the theoretical implications of the effects we observed: do our results speak to differences in how *much* change organizations make in response to social performance feedback, or to whether they make different *types* of organizational change ... or some combination of these?

While the research we present here is a step in improving how we conceptualize and operationalize social performance feedback, there are necessarily some challenges to building on the work we present here. An important first question for future work is one of where and how researchers should draw the boundaries of the firm's competitive group. The empirical context we examined provided an unequivocal and contextually defined reference group in which to examine our theoretical propositions, but future scholarship likely will have to determine the composition of the reference group *a priori*. Progress is already being made in this regard: Porac et al. (1999), Ketchen and Palmer (1999), Mezas et al. (2002), Baum et al. (2005), Massini et al. (2005), Labianca et al. (2009), and Hu et al. (2011) all suggest interesting ways to determine a meaningful social reference group that is smaller than the entire industry, and the extensive body of work by Fiegenbaum and colleagues is also informative (Fiegenbaum and Thomas 1990, 1995, Fiegenbaum et al. 1996, Shoham and Fiegenbaum 2002). The central issue in the construction of the social reference group is one of comparability among the included firms: social comparison occurs when actors view each other as commensurate rivals on a mutually relevant dimension



(Festinger 1954, Shibutani 1955, Goethals and Darley 1977). In simple industries, there may be only one competitive group comprising the entire industry. In more complex industries there may be many groups, each with top performance thresholds. Indeed, recent research suggests that firms may at any given time have multiple reference groups, depending upon the organizational goals they are trying to achieve (Labianca et al. 2009). In applying and extending the arguments and findings we generated in this paper, future researchers will have to make a determination on what firms are in the social reference group, so as to establish meaningful top performance and reference group thresholds.

A related complication for researchers is determining the marker of top performance to which group members make upward social comparisons (Festinger 1954). Organizational decision-makers within a nominally identical competitive group may hold differing perceptions of what constitutes “top” performance. In our setting there are a number of contextually derived possible top performance thresholds (Bundesliga Champion, Champions League, and UEFA Cup), and we proposed that the lowest of these, participating in the UEFA Cup, was the relevant threshold demarking the domain of top performance. To be sure, these empirical propositions are consistent with the economic benefits of achieving this level of performance in the Bundesliga (Barros, Ibrahim, and Symanski 2002). We also offered examples from another context (i.e., business schools) that lends itself to clear contextual determinants of different top performance thresholds. However, the exact specification of these thresholds in future research may require some context-specific exploratory research to establish the validity of different performance thresholds.

Researchers who want to operationalize our new HiBSAT measure will also need to ground aspects of that variable’s measurement in their specific context. Here, we proposed that HiBSAT updated once every year, and that there were periodic measurements of performance against that HiBSAT over the course of each year (i.e., after each round of play). This approach is certainly valid in our context, and may, in fact, apply to other contexts as well. For example, consider a firm that is ranked in the Fortune 100. While that ranking only updates once per year, between publications of the ranking the organization’s decision-makers are certainly comparing their market share and financial performance with

other firms in their reference group. Thus, it seems to us that the HiBSAT need not necessarily update on the same time interval as performance feedback measurement occurs. Nonetheless, this—like determination of the competitive reference group and the meaningful performance thresholds in that group—is ultimately a contextual question that future scholars will have to address in their respective studies.

## **Conclusion**

Insights from the BTF have found purchase with management scholars because, at their core, they have face validity: they seem accurately to describe the decision-makers we know and study. This paper has proposed that researchers can continue to make progress in this regard by revisiting one of the key drivers of organizational change articulated in the BTF—performance feedback based on social comparison. We have demonstrated that thresholds marking the lower boundary of the reference group as well as those marking top performance have salience to decision-makers who benchmark performance relative to a small competitive group. Moreover, our findings suggest that these decision-makers integrate knowledge of how their organization has performed in the past relative to these thresholds when contemplating organizational change. We hope additional work will continue to extend and refine BTF's core insights.

## REFERENCES

2006. Bundesliga Report: Deutsche Fußball Liga GmbH.
- Adner, R., C.E. Helfat. 2003. Corporate effects and dynamic managerial capabilities. *Strategic Management J.* **24** 1011-1025.
- Alesina, A., E. La Ferrara. 2002. Who trusts others? *Journal of Political Economy* **85** 207-234.
- Anderson, H.H., H.F. Brandt. 1939. Study of motivation involving self-announced goals of fifth grade children and the concept of level of aspiration. *Journal of Social Psychology* **10** 209-232.
- Audia, P.G., H.R. Greve. 2006. Less likely to fail: Low performance, firm size, and factory expansion in the shipbuilding industry. *Management Sci.* **52**(1) 83-94.
- Barros, C.P., M. Ibrahim, S. Symanski, eds. 2002. *The Comparative Economics of North American and European Sports*. Edward Elgar Publishing, Cheltenham, U.K.
- Baum, J.A.C., K.B. Dahlin. 2007. Aspiration performance and railroads' patterns of learning from train wrecks and crashes. *Organ. Sci.* **18**(3) 368-385.
- Baum, J.A.C., T.J. Rowley, A.V. Shipilov, Y.-T. Chuang. 2005. Dancing with strangers: Aspiration performance and the search for underwriting syndicate partners. *Admin. Sci. Quart.* **50**(4) 536-575.
- Beck, N., J. Brüderl, M. Woywode. 2008. Momentum or deceleration? Theoretical and methodological reflections on the analysis of organizational change. *Acad. Management J.* **51** 413-435.
- Beck, N., M. Meyer. 2012. Modeling team performance. *Empirical Economics* **43**(1) 335-356.
- Berman, S.L., J. Down, C.W.L. Hill. 2002. Tacit knowledge as a source of competitive advantage in the National Basketball Association. *Acad. Management J.* **45**(1) 13-31.
- Bothner, M.S., J. Kang, T.E. Stuart. 2007. Competitive crowding and risk taking in a tournament: Evidence from NASCAR racing. *Admin. Sci. Quart.* **52** 208-247.
- Boyle, E., Z. Shapira. 2012. The liability of leading: Battling aspiration and survival goals in the *Jeopardy!* Tournament of Champions. *Organ. Sci.* **23**(4) 1100-1113.
- Brenner, L., Y. Rottenstreich, S. Sood. 1999. Comparison, grouping, and preference. *Psychological Science* **10**(3) 225-229.
- Bromiley, P. 1991. Testing a causal model of corporate risk taking and performance. *Acad. Management J.* **34**(1) 37-59.
- Buunk, A.P., F.X. Gibbons. 2007. Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes* **102** 3-21.
- Chen, W.R., K.D. Miller. 2007. Situational and institutional determinants of firms' R&D search intensity. *Strategic Management J.* **28** 369-381.
- Cleaver, J. 2003. Lust for lists. *Workforce* **82**(5) 44.

- Connelly, B.L., L. Tihanyi, T.R. Crook, K.A. Gangloff. 2014. Tournament theory: Thirty years of contests and competitions. *J. Management* **40**(1) 16-47.
- Corley, K., D. Gioia. 2000. The rankings game: Managing business school reputation. *Corporate Reputation Review* **3**(4) 319-333.
- Cyert, R.M., J.G. March. 1963. *A Behavioral Theory of the Firm*, 2nd ed. Blackwell Publishers, Inc., Malden, Massachusetts.
- DiPrete, T.A., G.M. Eirich, M. Pittinsky. 2010. Compensation benchmarking, leapfrogs, and the surge in executive pay. *Amer. J. Sociology* **115**(6) 1671-1712.
- Elsbach, K.D., R.M. Kramer. 1996. Members' response to organizational identity threats: Encountering and countering the Business Week rankings. *Admin. Sci. Quart.* **41**(3) 442-476.
- Festinger, L. 1954. A theory of social comparison processes. *Human Relations* **7** 117-140.
- Fiegenbaum, A., S. Hart, D. Schendel. 1996. Strategic reference point theory. *Strategic Management J.* **17** 219-235.
- Fiegenbaum, A., H. Thomas. 1990. Strategic groups and performance: The U.S. insurance industry, 1970-84. *Strategic Management J.* **11** 197-215.
- Fiegenbaum, A., H. Thomas. 1995. Strategic groups as reference groups: Theory, modeling, and empirical examination of industry and competitive strategy. *Strategic Management J.* **16**(461-476).
- Fombrun, C.J., E.J. Zajac. 1987. Structural and perceptual influences on intraindustry stratification. *Acad. Management J.* **30**(1) 33-50.
- Garcia, S.M., A. Tor, R. Gonzalez. 2006. Ranks and rivals: A theory of competition. *Personality and Social Psychology Bulletin* **32**(7) 970-982.
- Gibson, C., F. Vermeulen. 2003. A healthy divide: Subgroup strength as a stimulus for team learning behavior. *Admin. Sci. Quart.* **48** 202-239.
- Gioia, D.A., K.G. Corley. 2002. Being good versus looking good: Business school rankings and the Circean transformation from substance to image. *Academy of Management Learning & Education* **1**(1) 107-120.
- Goethals, G.R., J. Darley. 1977. Social comparison theory: An attributional approach. J. Suls, R. L. Miller, eds. *Social Comparison Process: Theoretical and Empirical Perspectives*. Hemisphere, Washington, D.C.
- Gooding, R.Z., S. Goel, R.M. Wiseman. 1996. Fixed versus variable reference points in the risk-return relationship. *Journal of Economic Behavior and Organization* **29** 331-350.
- Greene, W.H. 2008. *Econometric Analysis*, 5th ed. Prentice-Hall, New Jersey.
- Greve, H.R. 1998. Performance, aspirations, and risky organizational change. *Admin. Sci. Quart.* **43**(1) 58-86.

- Greve, H.R. 2003a. A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Acad. Management J.* **46**(6) 685-702.
- Greve, H.R. 2003b. *Organizational Learning from Performance Feedback*. Cambridge University Press, Cambridge.
- Hammer, M. 1996. *Beyond Reengineering*. Harper Collins, New York.
- Heath, C., S. Huddart, M. Lang. 1999. Psychological factors and stock option exercise. *The Quarterly Journal of Economics* 601-627.
- Hoffman, P.J., L. Festinger, D.H. Lawrence. 1954. Tendencies toward group comparability in competitive bargaining. *Human Relations* **7** 141-159.
- Holcomb, T.R., R.M. Holmes, Jr., B.L. Connelly. 2009. Making the most of what you have: Managerial ability as a source of resource value creation. *Strategic Management J.* **30**(5) 457-485.
- Hu, S., D. Blettner, R.A. Bettis. 2011. Adaptive aspirations: Performance consequences of risk preferences at extremes and alternative reference groups. *Strategic Management J.* **32**(13) 1426-1436.
- Hunt, M.S. 1972. *Competition in the major home appliance industry, 1960-1970*. Unpublished Ph.D. Dissertation, Harvard University, Cambridge, MA.
- Isaac, M.S., R.M. Schindler. IN PRESS. The top-ten effect: Consumers' subjective categorization of ranked lists. *Journal of Consumer Research*.
- Jordan, A.H., P.G. Audia. 2012. Self-Enhancement and Learning from Performance Feedback. *Acad. Management Rev.* **37**(2) 211-231.
- Kelley, H.H. 1952. Two functions of reference groups. G. Swanson, T. Newcomb, E. Hartley, eds. *Readings in Social Psychology*. Society for the Psychological Study of Social Issues, New York, NY, 410-414.
- Ketchen, D.J., Jr., T.B. Palmer. 1999. Strategic responses to poor organizational performance: A test of competing perspectives. *J. Management* **25**(5) 683-706.
- Labianca, G., J.F. Fairbank, G. Andrevski, M. Parzen. 2009. Striving toward the future: Aspiration-performance discrepancies and planned organizational change. *Strategic Organization* (**7**) 433-466.
- Lant, T.K. 1992. Aspiration level adaptation: An empirical exploration. *Management Sci.* **38**(5) 623-644.
- Lehner, J. 2000. Shifts of reference points for framing of strategic decisions and changing risk-return associations. *Management Sci.* **46**(1) 63-76.
- Levinthal, D., J.G. March. 1981. A model of adaptive organizational search. *Journal of Economic Behavior & Organization* **2**(4) 307-333.
- March, J.G., Z. Shapira. 1987. Managerial perspectives on risk and risk taking. *Management Sci.* **33**(11) 1404-1418.

- March, J.G., Z. Shapira. 1992. Variable risk preferences and the focus of attention. *Psychological Rev.* **99**(1) 172-183.
- March, J.G., H.A. Simon. 1958. *Organizations*. John Wiley & Sons, Inc., New York.
- Martins, L.L. 2005. A model of the effects of reputational rankings on organizational change. *Organ. Sci.* **16**(6) 701-720.
- Massini, S., A.Y. Lewin, H.R. Greve. 2005. Innovators and imitators: Organizational reference groups and adoption of organizational routines. *Res. Policy* **34** 1550-1569.
- Medvec, V.H., S.F. Madey, T. Gilovich. 1995. When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology* **69**(4) 603-610.
- Mezias, S.J., Y.R. Chen, P.R. Murphy. 2002. Aspiration-level adaptation in an American financial services organization: A field study. *Management Sci.* **48**(10) 1285-1300.
- Miller, K.D., W.R. Chen. 2004. Variable organizational risk preferences: Test of the March-Shapira model. *Acad. Management J.* **47**(1) 105-115.
- Moliterno, T.P., C.M. Beckman. 2011. Behavioral and cognitive microfoundations of dynamic managerial capabilities: A multilevel structure.
- Moliterno, T.P., M.F. Wiersema. 2007. Firm performance, rent appropriation, and the strategic resource divestment capability. *Strategic Management J.* **28**(11) 1065-1087.
- Nyberg, A.J., Moliterno, T.P., Hale, D. & Lepak, D.P. (2014) "Resource-based Perspectives on Unit-level Human Capital: A Review and Integration." *Journal of Management* 40(1): 316-346.
- Ocasio, W. 1997. Towards an attention-based view of the firm. *Strategic Management J.* **18**(Special Summer Issue) 187-206.
- Olson, C.A., A. Schwab. 2000. The performance effects of human resource practices: The case of interclub networks in professional baseball, 1919-1940. *Industrial Relations* **39**(4) 553-577.
- Panagiotou, G. 2007. Reference theory: Strategic groups and competitive benchmarking. *Management Decision* **45**(10) 1595-1621.
- Pfeffer, J. 1985. Organizational Demography: Implications for Management. *California Management Rev.* **28**(1) 76-81.
- Ployhart, R.E., T.P. Moliterno. 2011. Emergence of the human capital resource: A multilevel model. *Acad. Management Rev.* **36**(1) 127-150.
- Porac, J.F., H. Thomas, C. Baden-Fuller. 1989. Competitive groups as cognitive communities: The case of the Scottish Knitwear Industry. *J. Management Studies* **26** 397-416.
- Porac, J.F., J.B. Wade, T.G. Pollock. 1999. Industry categories and the politics of the comparable firm in CEO compensation. *Admin. Sci. Quart.* **44** 112-144.

- Reger, R.K., A.S. Huff. 1993. Strategic groups: A cognitive perspective. *Strategic Management J.* **14** 103-124.
- Sauder, M., W.N. Espeland. 2009. The discipline of rankings: Tight coupling and organizational change. *Amer. Sociological Rev.* **74**(1) 63-82.
- Schimmer, M., M. Brauer. 2012. Firm performance and aspiration levels as determinants of a firm's strategic repositioning within strategic group structures *Strategic Organization* **10**(4) 406-435.
- Shibutani, T. 1955. Reference groups as perspectives. *Amer. J. Sociology* **60**(6) 562-569.
- Shinkle, G.A. 2012. Organizational aspirations, reference points, and goals: Building on the past and aiming for the future. *J. Management* **31**(1) 415-455.
- Shoham, A., A. Fiegenbaum. 2002. Competitive determinants of organisational risk-taking attitude: The role of strategic reference points. *Management Decision* **40**(2) 127-141.
- Short, J.C., T.B. Palmer. 2003. Organizational performance referents: An empirical examination of their content and influences. *Organizational Behavior and Human Decision Processes* **9** 209-224.
- Sirmon, D.G., S. Gove, M.A. Hitt. 2008. Resource management in dyadic competitive rivalry: The effects of resource bundling and deployment. *Acad. Management J.* **51**(5) 919-935.
- Sirmon, D.G., M.A. Hitt. 2009. Contingencies within dynamic managerial capabilities: Interdependent effects of resource investment and deployment on firm performance. *Strategic Management J.* **30**(13) 1375-1394.
- Sirmon, D.G., M.A. Hitt, R.D. Ireland. 2007. Managing firm resources in dynamic environments to create value: Looking inside the black box. *Acad. Management Rev.* **32**(1) 273-191.
- Staw, B.M., L.E. Sandelands, J.E. Dutton. 1981. Threat-rigidity effects in organizational behavior: A multilevel analysis. *Admin. Sci. Quart.* **26** 501-524.
- Washburn, M., P. Bromiley. 2012. Comparing aspiration models: The role of selective attention. *J. Management Studies* **49**(5) 896-917.
- Williams, K.Y., C.A.I. O'Reilly. 1998. Demography and diversity in organizations: A review of 40 years of research. In L. L. Cummings, B. M. Staw (Eds.), *Research in Organizational Behavior*, Vol. 20: 77-140. Greenwich, CT: JAI Press, Inc.
- Wiseman, R.M., P. Bromiley. 1996. Toward a model of risk in declining organizations: An empirical examination of risk, performance and decline. *Organ. Sci.* **7**(5) 524-543.
- Wood, J.V. 1989. Theory and research concerning social comparison of personal attributes. *Psychological Bulletin* **106**(2) 231-248.

**Table 1: Example of HiBSAT Calculations for the VFB Stuttgart Football Club in 1999/2000**

<b>Step 1 Calculate Historic Aspiration Level</b>	<b>Step 2 Determine HiBSAT</b>	<b>Step 3 Calculate Distance to HiBSAT</b>
<p>The VFB Stuttgart football club finished the 1998/99 season ranked 11<sup>th</sup>. Using this and other data for earlier years, we calculate their historical aspiration level (HAL) at 9.35 for 1999/2000 (see Eqn. 2). Thus, as a function of prior performance, the team would have historical aspirations of finishing the 1999/2000 season between the 9<sup>th</sup> and 10<sup>th</sup> rank.</p>	<p>Calculate the absolute value of the distance from the reference group threshold and the top performance threshold for VFB Stuttgart’s HAL of 9.35, recalling that top performance threshold is the 6<sup>th</sup> rank (i.e., the rank that allowed for UEFA cup participation) and the reference group threshold is the 15<sup>th</sup> rank:</p> <ul style="list-style-type: none"> <li>• Distance to top performance: <math> 9.35 - 6  = 3.35</math></li> <li>• Distance to relegation: <math> 9.35 - 15  = 5.65</math></li> </ul> <p>Since the absolute value of the distance to the UEFA Cup is the smaller value, we set VFB Stuttgart’s HiBSAT for 1999/2000 at the top performance threshold. We use this as the HiBSAT for the whole of the 1999/2000 season.</p>	<p>Measure VFB Stuttgart’s performance relative to HiBSAT after each round in the 1999/2000 season by making comparisons, in points, to the teams that occupy the top performance threshold (i.e., the team’s HiBSAT). For example after the 15<sup>th</sup> round of the 1999/2000 season, Werder Bermen was on the 6<sup>th</sup> rank with 23 points, while VFB Stuttgart had 21 points, placing them two points below the performance needed to be on their HiBSAT. Thus we enter -2 as VFB Stuttgart’s Performance relative to HiBSAT this period.</p>



**Table 2: Descriptive Statistics and Correlation Table<sup>1</sup>**

	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13
1 N new players	1.12	1.07	1.00												
2 Distance from ref. group threshold	5.51	7.46	-0.02	1.00											
3 Distance from top performance threshold	-3.93	7.33	-0.06	0.77	1.00										
4 Perf. below HiBSAT	-2.07	3.86	-0.10	0.37	0.54	1.00									
5 Perf. above HiBSAT	2.52	4.36	-0.06	0.66	0.54	0.32	1.00								
6 HiBSAT avoid relegation	0.47	0.50	-0.08	0.33	-0.33	0.33	0.22	1.00							
7 Round	18.00	9.52	0.02	0.31	-0.23	-0.23	0.23	-0.01	1.00						
8 Round squared	414.67	352.25	0.00	0.29	-0.22	-0.22	0.21	-0.01	0.97	1.00					
9 Moving average heterogeneity	0.32	0.04	0.09	0.19	0.26	-0.02	0.01	-0.37	-0.21	-0.21	1.00				
10 Engagement in Europ. Cup	0.22	0.41	0.10	0.14	0.29	0.04	-0.05	-0.45	-0.27	-0.26	0.27	1.00			
11 New coach	0.01	0.12	0.10	0.07	-0.09	-0.10	-0.06	-0.01	0.03	0.02	0.01	-0.02	1.00		
12 Tenure of new coach	1.84	6.00	0.00	0.08	-0.27	-0.17	-0.07	0.04	0.30	0.32	-0.02	-0.11	0.02	1.00	
13 Difference in ranks	0.00	7.43	0.05	0.57	-0.58	-0.31	-0.37	0.26	0.00	0.00	-0.18	-0.19	0.07	0.11	1.00

<sup>1</sup> N=6534, pairwise correlations >|0.02| significant at  $p < 0.05$  level

**Table 3: Fixed-effect GEE Poisson-Models of Initial Lineup Changes<sup>1</sup>**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Distance to top performance threshold			-0.0133** (0.0026)	-0.0127* (0.0059)		
Distance to reference group threshold		-0.0123** (0.0026)		-0.0007 (0.0060)		
Performance below HiBSAT					-0.0185** (0.0040)	
Performance above HiBSAT					-0.0126** (0.0040)	
HiBSAT top perf.: distance to top						-0.0169* (0.0066)
HiBSAT top perf.: distance to ref.						0.0052 (0.0062)
HiBSAT avoid rel.: distance to top						-0.0070 (0.0066)
HiBSAT avoid rel.: distance to ref.						-0.0160* (0.0075)
HiBSAT avoid relegation	-0.0989** (0.0362)	-0.0930* (0.0364)	-0.0934* (0.0363)	-0.0933* (0.0363)	0.0680 (0.0462)	0.0618 (0.0657)
Round	0.0334** (0.0058)	0.0380** (0.0059)	0.0299** (0.0058)	0.0303** (0.0069)	0.0330** (0.0058)	0.0306** (0.0069)
Round squared	-0.0008** (0.0002)	-0.0008** (0.0002)	-0.0007** (0.0002)	-0.0007** (0.0002)	-0.0007** (0.0002)	-0.0007** (0.0002)
Moving average heterogeneity	1.6506** (0.4947)	1.6147** (0.4957)	1.6519** (0.4957)	1.6499** (0.4960)	1.5788** (0.4962)	1.6760** (0.4964)
Engagement in European Cup	0.1922** (0.0367)	0.1839** (0.0369)	0.1809** (0.0369)	0.1810** (0.0369)	0.2065** (0.0369)	0.2013** (0.0378)
New coach	0.5454** (0.0766)	0.5049** (0.0770)	0.5038** (0.0769)	0.5034** (0.0770)	0.4895** (0.0771)	0.4906** (0.0770)
Tenure of new coach	-0.0016 (0.0023)	-0.0042† (0.0024)	-0.0047† (0.0024)	-0.0047† (0.0024)	-0.0045† (0.0024)	-0.0050* (0.0024)
Difference in ranks of teams playing	0.0103** (0.0018)	0.0060** (0.0020)	0.0056** (0.0020)	0.0056** (0.0020)	0.0052** (0.0020)	0.0050* (0.0020)
Constant	-0.7242** (0.1781)	-0.7027** (0.1786)	-0.6802** (0.1787)	-0.6810** (0.1788)	-0.7413** (0.1788)	-0.7513** (0.1801)
$\chi^2$ (DF)	365.6 (49)	384.6 (50)	389.7 (50)	389.7 (51)	400.4 (51)	402.6 (53)

<sup>1</sup>N=6534, standard errors in parentheses, AR(1) error structure, season dummy variables included in all models

\*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.10$

**Table 4: Fixed-effect Logit Models of Coach Change<sup>1</sup>**

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Distance to top performance threshold			-0.1614** (0.0278)	-0.0048 (0.0589)		
Distance to reference group threshold		-0.1952** (0.0315)		-0.1907** (0.0621)		
Performance below HiBSAT					-0.1754** (0.0336)	
Performance above HiBSAT					-0.3141* (0.1479)	
HiBSAT top perf.: distance to top						-0.0927 (0.0632)
HiBSAT top perf.: distance to ref.						-0.0884 (0.0551)
HiBSAT avoid rel.: distance to top						0.0787 (0.0680)
HiBSAT avoid rel.: distance to ref.						-0.3630** (0.0804)
HiBSAT avoid relegation	-0.1562 (0.2497)	-0.0778 (0.3459)	-0.1369 (0.3245)	-0.0795 (0.3451)	2.0866** (0.4763)	1.7882** (0.5300)
Round	0.1921** (0.0707)	0.2571** (0.0701)	0.1362† (0.0726)	0.2537** (0.0865)	0.1970** (0.0694)	0.2526** (0.0833)
Round squared	-0.0043** (0.0016)	-0.0051** (0.0016)	-0.0038* (0.0016)	-0.0051** (0.0017)	-0.0045** (0.0016)	-0.0052** (0.0016)
Engagement in European Cup	-0.4347 (0.2741)	-0.6526† (0.3415)	-0.6410† (0.3322)	-0.6536† (0.3398)	-0.3885 (0.3259)	-0.4249 (0.3190)
Tenure of new coach	-0.0203 (0.0186)	-0.0684* (0.0310)	-0.0624* (0.0271)	-0.0685* (0.0303)	-0.0648* (0.0257)	-0.0704* (0.0306)
Difference in ranks of playing teams	0.1013** (0.0187)	0.0322† (0.0193)	0.0419* (0.0195)	0.0321† (0.0193)	0.0281 (0.0187)	0.0286 (0.0179)
Log-likelihood	-386.1	-360.8	-365.7	-360.8	-356.0	-354.0
$\chi^2$ (DF)	317.3 (16)	577.6 (17)	685.6 (17)	670.1 (18)	3271 (18)	2575 (20)

<sup>1</sup>N=5874, robust standard errors in parentheses, season dummy variables included in all models

\*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.10$