CREMA

Center for Research in Economics, Management and the Arts

Optimistic and Positivity Biases in Employee Ratings: Empirical Evidence from Professional Soccer

Working Paper No. 2015-02

CREMA Südstrasse 11 CH - 8008 Zürich www.crema-research.ch

Optimistic and Positivity Biases in Employee Ratings:

Empirical Evidence from Professional Soccer

Steffen Merkel

EBS Universität für Wirtschaft und Recht

Sascha L. Schmidt

WHU – Otto Beisheim School of Management and

CREMA – Center for Research in Economics, Management and the Arts

Benno Torgler

Queensland University of Technology and

CREMA – Center for Research in Economics, Management and the Arts

Author Note

Steffen Merkel, Center for Sports Marketing, EBS Universität für Wirtschaft und Recht, Oestrich-Winkel, Germany; Sascha L. Schmidt, Center for Sports and Management, WHU – Otto Beisheim School of Management, Düsseldorf, Germany and CREMA – Center for Research in Economics, Management and the Arts, Südstrasse 11, Zürich, Switzerland; Benno Torgler, School of Economics and Finance, Queensland University of Technology, Brisbane, Australia and CREMA – Center for Research in Economics, Management and the Arts, Südstrasse 11, Zürich, Switzerland.

Correspondence concerning this article should be addressed to Steffen Merkel, Center for Sports Marketing, EBS Business School, EBS Universität für Wirtschaft und Recht, Rheingaustrasse 1, 65375 Oestrich-Winkel, Germany, Phone: +49 611 7102 1442, Fax: +49 611 7102 10 1442, E-Mail: steffen.merkel@ebs.edu.

Abstract

This study uses a real case from professional soccer to examine intertemporal rating errors in performance appraisals. Motivated by research that extends the (prospective) optimistic bias and (retrospective) positivity bias to others, we formulate five hypotheses on the reflection of these biases in employee ratings and on rater-/ratee-related moderators of effect sizes. Employing unique assessment data for 164 players from a German Bundesliga club's youth academy, we show that the ratings of *predicted* and *remembered* performance are indeed higher than the talents' *actual* performance throughout a season. The differences depend positively on both the rater's experience and the amount of information available on the ratee but are not significantly influenced by the closeness of their relationship. Moreover, the (prospective) anticipation is even more positively biased than the (retrospective) recollection of the players' performance, which highlights an interesting asymmetry between looking forward and backward. Because the academy's appraisal and promotion schemes resemble typical company practices, we subsequently transpose our new understanding of rating errors – especially their intertemporal nature – to the business domain, suggesting practical implications for the design of corporate appraisal systems.

Keywords: optimistic bias; positivity bias; rating error; performance appraisal; professional soccer

Performance appraisal systems serve numerous important purposes within organizations; for example, as a basis for promotion and merit pay assignment or as indicators of development needs (Cleveland, Murphy, & Williams, 1989). Given their importance, research evidence of frequent systematic errors in supervisor ratings – including the halo (Balzer & Sulsky, 1992; Cooper, 1981), leniency, centrality (Bol, 2011; Ng et al., 2011) and similar-to-me effects (Antonioni & Park, 2001; Latham & Wexley, 1982) and race, gender, and age biases (McShane, 1990; Stauffer & Buckley, 2005) – imply significant costs for the organization (Heneman, Moore, & Wexley, 1987).

One potential source of such rating bias is timing: supervisors usually monitor their subordinates' *actual* performance over a certain period but provide a rating of their *remembered* performance at period end, sometimes accompanied by an assessment of the employee's *predicted* potential (Rothwell, 2010). Hence, although a rating should ideally be the average of many instantly reported moment-by-moment evaluations (Kahneman, 2003; Kahneman, Wakker, & Sarin, 1997), in practice conscious memory cannot achieve such a temporally integrated performance record. Summary ratings thus tend to include less effortful heuristics and biases and are a fallible estimate of true performance (Kahneman, 2003; Reb & Cropanzano, 2007).

In particular, human prospective and retrospective judgments frequently include two types of cognitive error: an optimistic bias, the tendency to hold unrealistically optimistic or illusory views about the future (Taylor & Brown, 1988), and/or a positivity bias, a propensity to see the past through "rose-colored glasses" (Skowronski, 2011). Put differently, both biases describe an upward distortion of human judgment – either in looking forward (the optimistic bias) or in looking backward (the positivity bias). Not only does the literature extensively document such future- and past-oriented distortions in self-relevant autobiographical events (Sedikides & Gregg, 2007; Taylor & Brown 1988), but laboratory-

based research indicates that unrealistic optimism (see Regan, Synder, & Kassin, 1995) and retrospective positivity through temporal self-appraisal (see Konrath & Ross, 2003) also pervade our thoughts about others. Such empirical findings suggest that we may hold "rosy views" on the future and past of others in real-life situations, which manifest as rosy forward- and backward-looking evaluations in the most common organizational rating scenario (Bernardin & Villanova, 1986), supervisory appraisal of employees' past performance and future potential.

In this paper, we search for signs of an optimistic and/or positivity bias in employee ratings by analyzing unique empirical data: the scores from semi-annual performance appraisals in the youth academy of a German Bundesliga club. These young players receive three types of evaluations from their coaches: a global rating of their predicted future performance, a global rating of their remembered performance during the last half-year, and a record of instantly reported (i.e., temporally unbiased[1]) ratings of their actual performance in individual matches. These equally scaled ratings emphasize development needs, provide feedback, and support promotions to the next age group, purposes that are not only common in a business context but show striking parallels with up-or-out systems in professional service firms (see Ghosh & Waldmann, 2010). Our results should therefore be generalizable.

To increase understanding of rating errors, particularly their intertemporal nature, this study leverages the fact that youth academies provide a controlled setting in which all "employees" have a relatively homogeneous job profile and share the objective of becoming a professional player (Jung, Schmidt, & Torgler, 2012). The analysis is novel in that although prior research examines the impact of serial position effects[2] (Steiner & Rain, 1989) or dynamic performance trends and variations (Reb & Cropanzano, 2010) on employee

[1]Hereafter, we refer to an actual in the sense of a temporally unbiased performance rating – which can, however, still be subject to static judgmental errors.

[2] Studies on serial position effects examine if the position of a single good/poor performance in a sequence of average performances affects overall performance ratings.

evaluations, to the best of our knowledge, no direct comparison of assessments provided at different points in time has as yet been undertaken. Our study of predicted, remembered, and actual "on-line" performance ratings is moreover a real-life example of two psychological biases that have predominantly been observed in laboratory settings before. This innovative research setup does not only allow us to determine if optimistic and positivity biases are reflected in performance ratings *at all*, but even if there is an asymmetry between the two.

The article proceeds as follows: after theoretically deriving five hypotheses on the main effects and potential moderators of the effect size, we describe our sample and method and then report our bi- and multivariate results. We conclude by discussing the practical implications and limitations of the findings and suggesting directions for future research.

Theory and Hypotheses

The Optimistic Bias and Positivity Bias in Self-Relevant Autobiographical Events

The tendency to conduct research on the optimistic bias and positivity bias primarily in the context of self-relevant autobiographical events stems from two factors: First, most thoughts about the future and the past revolve around the self (Johnson & Sherman, 1990). Second, both biases can be linked to contemporary theories about the self, particular those involving self-enhancement (Chambers & Windschitl, 2004; Sedikides & Green, 2009; Skowronski, 2011). In particular, human cognition tends to preferentially remember and process all information affirming a favorable self-image, including positive recollections of past actions and optimistic expectations for the future (Sedikides & Strube, 1997; Skowronski, 2011). Not only are the resulting self-serving biases such as unrealistic optimism and retrospective positivity pervasive, enduring, and systematic, but illusory self-conceptions are actually important for the mental health and well-being of normal personalities (Taylor & Brown, 1988).

The forward-looking optimistic bias and the backward-looking positivity bias enter autobiographical expectations and memories through reconstructive processes: humans construct (i.e., create) and reconstruct (i.e., distort) the past and future in the present by means of coherent narratives that anticipate and reflect their own experiences (Baumeister & Newman, 1994; Johnson & Sherman, 1990). Because this process does not strictly follow patterns of formalized abstract thinking, it allows biases in the form of affective evaluations to come into play (Mitchell, Thompson, Peterson, & Cronk, 1997). In other words, humans "construct stories that bolster their sense of self-worth" (Baumeister & Newman, 1994, p. 686) and have an ego that "fabricates and revises history" (Greenwald, 1980, p. 603).

The *optimistic bias* is commonly described as a tendency to consider oneself less likely than others to experience negative events and more likely to experience positive events. As the definition suggests, it is usually operationalized via comparative judgments of self-versus-other risk regarding the occurrence of a given future event (Helweg-Larsen & Shepperd, 2001). Such relative measures are also used to determine unrealistic or comparative optimism (Weinstein, 1980) and illusions of unique invulnerability (Perloff & Fetzer, 1986), which essentially describe the same phenomenon (Shepperd, Carroll, Grace, & Terry, 2002). When personal risk dispositions are unknown, the optimistic bias is difficult to verify on an individual level. That is, individuals' generally hopeful and confident thoughts about the future (Taylor & Brown, 1988; Tiger, 1979) may actually be justified if optimistic subjects have favorable prospects. On a group level, however, systematic unrealistic optimism is easily detectable if all subjects consider their chances of experiencing a negative event as below average when logically, not everyone can be better off than an average group member (Weinstein, 1980). The optimistic bias is even documented for numerous partly or entirely uncontrollable events like unwanted pregnancies (Burger & Burns, 1988), car accidents (McKenna, 1993), natural disasters (Burger & Palmer, 1992), and purely chance-determined

lotteries (Langer, 1975). These observations clearly demonstrate that, although humans are realistic in their evaluations of an average other, their judgments of their own futures are positively distorted (Shepperd et al., 2002).

Such positive distortion also characterizes personal recollections of the past: because the cognitive system recalls and processes favorable personality information more efficiently than unfavorable personality information, memories are often rose colored. This *positivity bias,* which is connected to a hedonic valence effect in autobiographical memory and operates through several mechanisms (Skowronski, 2011; Taylor & Brown, 1988), implies that individuals remember the past better than it actually was. On the one hand, they selectively forget unfavorable content, remembering pleasant events slightly better than unpleasant events because negative affect intensity fades faster than positive affect intensity (Walker, Skowronski, & Thompson, 2003; Walker, Vogl, & Thompson, 1997). This latter apparently stems from the ability to mobilize cognitive, physiological, emotional, and social resources after an adverse event to cope with its immediate consequences and minimize a long-term emotional burden (Taylor, 1991). The net result is a relative preservation of positive over negative emotion. On the other hand, they modify recalled content; for instance, by positively inflating recollections of their high school grades (Bahrick, Hall, & Berger, 1996) or cardiovascular risk categories (Croyle et al., 2006). Even when they do recall content accurately, they may disassociate themselves from it; for example, by distancing themselves more from former selves with unpleasant implications for the current self-view than from those with pleasant implications (Wilson & Ross, 2000), thereby positioning successes nearer and failures further away (Ross & Wilson, 2002).

Whereas the above definitions, mechanisms, and causes of the optimistic and positivity biases clearly link both phenomena to the self, empirical evidence further suggests that self-relevant cognitive biases in general (e.g., Aron & Aron, 1986) and optimistic (e.g., Regan et

al., 1995) and positivity biases (e.g., White, Coppola, & Multunas, 2008) in particular extend to our thoughts about others. Such evidence would explain theoretically why these two primarily self-relevant biases are incorporated into the real-life performance ratings of others.

Extensions of Self-Relevant Cognitive Biases to Others

The most thorough explanation of how self-conceptions, and thus self-relevant cognitive biases, can be extended to others is provided by the self-expansion model, which postulates that, when forming relationships, humans include resources, perspectives, and identities of close others in the self to increase their own potential efficacy. In short, we perceive aspects of the other – whether family member, spouse, roommate, or colleague – as partly our own (Aron & Aron, 1986; Aron, Lewandowski, Mashek, & Aron, 2013; Aron, Mashek, & Aron, 2004). Analogously, this perception can be extended to coaches and their young protégés because the professional club environment requires the two to collaborate as closely as a company supervisor-subordinate dyad. Thus, for example, a typical "soccer work week" for the Under 19 squad comprises 1–2 training sessions per day, joint team meals, appointments with physiotherapists or psychologists, and 1–2 matches, which can easily total 40 hours per week. Moreover, in both the supervisor-subordinate and coach-player relationships, the protégé's rating at least partly mirrors the supervisor's own managerial capabilities (Longenecker, Sims, & Gioia, 1987), exemplifying a central feature of personal closeness: the fact that "one experiences the other's outcomes to some extent as one's own" (Aron et al., 2013, p. 106).

When experiencing and appreciating the world from the other's perspective, we extend our self-related attributional and cognitive biases to evaluations of the other person so that self- and other-judgments become increasingly similar (Aron et al., 2004). For example, self-reference effects (i.e., the better recall of self-related than other-related information) vanish (Symons & Johnson, 1997). Individuals also attribute the behavior of close others to

situational rather than dispositional factors, which resembles interpretations of our own behavior (Aron, Aron, Tudor, & Nelson, 1991; Sande, Goethals, & Radloff, 1988). This perspectives dimension of self-expansion is particularly relevant to our investigation.

Extensions of the Optimistic Bias to Others

Besides demonstrating this extension of cognitive biases to others, research also shows that the tendency applies particularly to the optimistic bias, as indicated by the fact that the level of unrealistic optimism about one's own future depends on a comparative target. When this target "average other" is distant, dissimilar, and vague, individuals are confident in positively comparing their risk of victimization (Helweg-Larsen & Shepperd, 2001). This confidence decreases, however, in comparisons with close, comparable, and specific targets such as a good friend (Helweg-Larsen & Shepperd, 2001), which a self-expansionist perspective might attribute to overoptimism for someone close. Another possibility, however, is that the lower comparative optimism results from individuation of the target; that is, a good friend is not only closer to the self than the "average other" but, unlike abstract stereotypes like "same-sex peer" or "typical employee," is an individual human being. Accordingly, person positivity (Sears, 1983), the heuristic to value *any* individual social object more favorably than abstract stereotypes, might explain lower personal optimism in comparisons with a close friend (Drake, 1984).

This explanation is refuted, however, by the second study in Regan et al. (1995), which asks subjects to estimate the chances that they, a close friend, a mere acquaintance, or a complete stranger will experience a number of positive and negative future life events. The authors identify a descending order of unrealistic optimism that cannot be attributed to the abstract group-like nature of any of the four individual comparison targets. They thus conclude that "the umbrella of optimistic biases with which we shield ourselves from the harsh drops of cold reality may not cover all people, but certainly seems large enough for

ourselves and those to whom we feel close" (p. 1080). This finding supports the assumption that humans extend their optimistic bias to others and behave as if the other's future outcomes were partly their own. The view is also supported by Helweg-Larsen and Shepperd's (2001) claim that psychological closeness moderates the level of unrealistic optimism we hold for others.

Such psychological closeness can even trigger comparative optimism for impersonal objects, as when technical experts believe that their own specializations will become more innovative than other fields. Foresight scholars attribute this insider bias to high emotional investment and a strong attachment to future outcomes in one's own domain (Tichy, 2004). In fact, experts are especially prone to commit cognitive errors in forecasting if the rated object is dynamic and the stimuli involve human behavior (Zaleskiewicz, 2011), which is obviously the case when predicting an employee's future performance. In line with these theoretical considerations, we hypothesize the following:

Hypothesis 1: Ratings of an employee's predicted future performance are optimistically biased and thus higher than ratings of his/her actual current performance.

Extensions of the Positivity Bias to Others

Although in investigating the optimistic bias, retrospective positivity research tends to concentrate on self-relevant autobiographical events, several studies do indicate that we see not only our own past but also that of others through rose-colored glasses. In the words of White et al. (2008), "self-enhancing memory distortions suggest that special memory processes are reserved for information about ourselves [....] However, the research on memories of others has resulted in theories that the same basic processes can account for autobiographical memories in general" (p. 293). These authors specifically refer to life scripts (Berntsen & Rubin, 2004; Rubin & Berntsen, 2003), which picture the life of an

idealized average person as comprising mostly desirable information, which can account for positive distortions in the encoding and retrieval of autobiographical events.

Accordingly, humans may brighten their memories of others, such as when widows idealize their dead husbands (Woodfield & Viney, 1984). Similarly, among living subjects, individuals may evaluate their friends as positively as themselves. Apparently, because we ourselves affectively respond to pleasant feedback, we in turn provide others with equally "accurate" information to match their own inflated self-concept (White et al., 2008). A retrospective positivity bias can even emerge if the remembered object is impersonal, such as the universal human tendency to use and recall evaluatively positive words more frequently, facilely, and diversely than evaluatively negative words (see the early work on the Pollyanna principal by Boucher & Osgood, 1969; and Matlin & Stang, 1978). We draw on the above evidence for a self-related positivity bias and its extension to judgments of others to formulate our second hypothesis:

Hypothesis 2: Ratings of an employee's remembered past performance are positively biased and thus higher than ratings of his/her actual performance during that period.

Moderators of the Optimistic Bias and the Positivity Bias as Extended to Others

According to the literature, if our empirical results confirm the hypothesized prospective optimism (H_1) and retrospective positivity (H_2) in employee ratings, these biases could be moderated by characteristics of the ratee and rater and their relation to each other:

Amount of information on the ratee. The amount of information about a rated object – whether the self or another person – can moderate unrealistic optimism. In particular, individuals' illusory expectations for their own future life events arise from the richly differentiated knowledge that their memories store about self. Then, to rationalize assessments that are in fact overly optimistic, they selectively (ab)use such information by focusing on factors that positively affect their chances (e.g., personal risk behavior, prior

experience, family history) while neglecting detrimental elements (Regan et al., 1995; Shepperd et al., 2002). The same mechanism applies to judgments of others; individuals store similarly favorable facts on the precautious behavior and personal attributes of family members or friends (Helweg-Larsen & Shepperd, 2001), and the more information available, the higher the optimistic bias. This observation is consistent with the finding that subjects' optimism for their own future decreases once specific personalized details are provided on a comparison target (Weinstein, 1980; Weinstein & Lachendro, 1982).

A similar argument can be made for the relation between the amount of available information and retrospective positivity. Because the cognitive system recalls and processes pleasant information more efficiently than unpleasant information and negative event-prompted emotions fade faster than positive ones (Skowronski, 2011; Walker & Skowronski, 2009), the more that is known about a ratee's past – assuming a constant ratio of positive to negative information – the higher the net preserved amount of positive emotions. We thus hypothesize the following:

Hypothesis 3: The more information available on a ratee, the more optimistically biased the prospective ratings and the more positively biased the retrospective ratings.

Rater experience. Although rater experience can moderate rater optimism in future-oriented judgments, experiencing a negative event does not increase confidence that "lightning never strikes twice" (Helweg-Larsen & Shepperd, 2001, p. 86). In fact, experienced subjects are less optimistically biased because they are less illusory about their control over future outcomes, can more easily imagine themselves in the role of victim, and overestimate the baseline risks of events encountered previously (Helweg-Larsen & Shepperd, 2001; Weinstein, 1989). Although this negative effect of experience on optimism is mainly documented for self-relevant events, there is no reason it should not extend to evaluations of others as in the case of prospective employee ratings.

Retrospective positivity can also be moderated by experience, as shown by Spence and Keeping's (2010) finding that in assessing employee performance, adept supervisors map new stimuli onto a richer base of expert knowledge than novice raters. Hence, ceteris paribus, the more appraisal experience supervisors have, the less lenient should be their global retrospective performance ratings of employees:

Hypothesis 4: The more experienced a rater, the less optimistically biased the prospective ratings and the less positively biased the retrospective ratings.

Closeness of the ratee/rater relationship. As already discussed, psychological closeness has such moderating effects as successively lowering unrealistic optimism for self, friend, acquaintance, and then stranger (Regan et al., 1995) and increasing comparative optimism with growing proximity and equation of the close one's prospects with one's own. Yet the level of retrospective positivity in judgments on others can also be increased by the closeness of the relationship between rater and ratee. For example, in a study of whether subjects favorably distort personal memories through temporal self-appraisal (Ross & Wilson, 2002) when evaluating others, Konrath and Ross (2003) find that the tendency to perceive successes as recent and failures as distant may also bias our assessments of close others. This finding suggests that retrospective positivity may increase with the closeness of the rater/ratee relationship:

Hypothesis 5: The closer the rater-ratee relationship, the more optimistically biased the prospective ratings and the more positively biased the retrospective ratings.

Data and Methods

Participants

To assess the role of the above biases in employee ratings, we investigate the performance ratings of players (analogous to employees) from the youth academy of one of the 36 professional soccer clubs in the German Bundesliga's first and second division.

Because these mandatory externally audited academies (see van Hoecke et al., 2011) must meet very rigid standards (Deutscher Fußball-Bund, 2010), they are recognized as the country's primary source of elite talent (Jung et al., 2012) and a key reason for Germany's current success in professional soccer (Deutsche Fußball Liga, 2010). Hence, our subjects, all males between 10 and 19 years old (M = 14.2), are among the best young soccer players in Germany, although 26% of them have foreign citizenship.

Players in the academies are divided into eight teams – the Under 12, -13, -14, -15, -16, -17, -19, and -23 squads – and undergo semi-annual performance appraisals. We analyze ratings for the 2010/11, 2011/12, and 2012/13 seasons (i.e., for three years). The time span for each observation is one half-year. Between 2010/11 and 2012/13, 274 youngsters played in the academy for a total of 973 half-years. However, a subject might have an incomplete evaluation record for a period if he entered the academy very recently, played only few games, or was injured. For example, 263 half-year observations did not contain any prospective and/or retrospective ratings of predicted and/or remembered performance. 345 additional half-year records of performance ratings from individual matches comprised an unrepresentatively low number of entries.[3] This total of 608 observations was not included in the analysis. Accordingly, the final sample contained 365 half-year observations from 164 young players.

Procedure

The youth academy's appraisal process, from which our secondary empirical data are taken, is as follows: For every half-year, lasting either from July until December or from January until June, coaches perform three types of evaluations. During the half-year itself,

[3]To ensure a meaningful comparison of a player's predicted, remembered, and actual performance ratings, the record of the latter must integrate a sufficient number of games. We thus include youngsters in the sample if they receive evaluations in at least six regular season games during the focal half-year, which corresponds to 46% to 75% of their team's matches, depending on the respective age group. Our results are also robust, however, to the specification of alternative thresholds of between five and seven games. Moreover, (non)parametric tests of the hypotheses show that coaches do not assign ratings after particularly successful or poor games, suggesting that the match outcomes in our sample are representative of "average" games.

they rate their players' performance in individual regular season games on a Likert scale from 1 (very low) to 10 (very high). Because these games constitute *the* most important events and grades are reported without delay, the averages of these ratings approximate the athletes' actual performance over that period. After the half-year, coaches also assign a retrospective summary rating each player's "remembered" performance during the last half-year and a prospective summary rating his "predicted" performance in the future. Both these are also scaled from 1 to 10. The evaluations are then discussed in a feedback session between the coach and the player.

After the second half-year in June, the coaches decide whether a youngster will be promoted to the next age group or must leave the academy[4] based on the impressions reflected in the ratings, which are stored on the club's Intranet and accessible to academy staff, scouts, and coaches. Because coaches are responsible only for one age group and do not move up with a cohort, most athletes get a new coach when the new season starts in July. The only exceptions are the Under 19 and Under 23 squads, in which players can remain for more than one season. Because of personnel turnover, the eight teams studied were trained by 15 different coaches during the 3-year investigative period.

Measures

Our two main effects of interest, the optimistic bias (H_1) and the positivity bias (H_2), act as dependent variables in our multivariate regressions. The independent variables comprise three measures for the hypothesized moderators of the optimistic and the positivity bias (H_3-H_5), control variables that might additionally influence the coaches' ratings, and categorical semester dummies. Table 1 lists the summary statistics for all the variables, which are separately described below.

--- Insert Table 1 about here ---

[4] Between the 2010/11 and 2012/13 seasons, 117 (42.7%) of the 274 players had to leave the academy. Another 25 (9.1%) quit voluntarily. That is, more than 50% of the subjects dropped out during the 3-year observation period.

Optimistic bias. In line with the common practice of measuring optimism by comparing expectations with reality (Armor & Taylor, 1998), we operationalize optimistic bias as the difference between a player's *predicted* performance and his *actual* performance during the half-year under investigation. Logically, some players are likely to improve, some to stagnate, and some to deteriorate, meaning that not everybody will be better off. The longitudinal analysis depicted by Figure 1 does indeed confirm that actual performance ratings for our sample do *not* trend upward significantly during the observation period. Hence, a consistently positive deviation of predicted from actual performance evaluations would indicate a systematic optimistic bias in forward-looking employee assessments.

--- Insert Figure 1 about here ---

Positivity bias. Following studies that treat retrospective positivity as a modification rather than a loss of memory content, we measure positivity bias as the difference between a player's *remembered* performance and *actual* performance during the half-year under investigation. We anticipate that, as reported in Bahrick et al. (1996) and Crary (1966), individuals may recall past achievements as being better than they actually were.

Amount of information on the ratee. The amount of available information on a player is represented by the variable *Ratee Age* (in years), which we choose instead of, for instance, tenure because the youth academy begins acquiring knowledge on a talent long before he enters the club. The club's eight salaried and additional freelance scouts, particularly, systematically compile strengths, weaknesses, and performance statistics on candidates at an early stage to substantiate admission decisions.

Rater experience. Because temporal measures are frequently used to operationalize rater experience (e.g., Schuh, 1973; Spence & Keeping, 2010), we capture this aspect with the variable *Rater Tenure*, which equals the number of months that a coach has worked at the academy (by the end of the half-year). We select tenure because the academy's IT-based,

multi-dimensional, and data-backed appraisal process requires very specific rater experience as it strongly differs from less structured practices in other clubs. Christensen (2009), for instance, portrays a more usual soccer talent identification approach where coaches rely on their practical sense, feel for the game, and implicit classificatory schemes.

Closeness of the rater/ratee relationship. Consistent with studies that define rater-ratee familiarity by the length of the performance-relevant contact (Jacobs & Kozlowski, 1985), the variable *Relationship Closeness* corresponds to the number of months that a player has already been trained by his current coach (at the end of the half-year).

Control variables. We include several soccer- or academy-specific control variables that might additionally influence the coaches' ratings. First, because performance determinants underlying the ratings vary by positional role (Di Salvo et al., 2007), a player's position is captured by binary dummy variables for *Goalkeeper*, *Defender*, *Striker*, and *Flexible Position* (i.e., *Midfielder* represents the reference category). Second, the variable *Relative Age* reflects the calendar date of a subject's birthday and can take values between 1 and 365. There is strong evidence that older talents who are born shortly after a cutoff date are selected and evaluated favorably (Helsen, van Winckel, & Williams, 2005; Musch & Grondin, 2001). In this case, the cutoff date is January 1st – lower values of *Relative Age* thus imply a higher age and should consequently correspond to higher ratings. Third, given previous documentation of ratings being affected by a ratee's racial or ethnic origin (Kraiger & Ford, 1985; Landy & Farr, 1980), the dummy variable *Foreign* equals 1 if a player is a non-German citizen. Fourth, in recognition of the recency effect by which the most recent impressions are better recalled (Murdock, 1962) and can disproportionately influence global performance ratings (Steiner & Rain, 1989), the variable *Last Game* is the deviation of a subject's last performance from his average grade in individual games during the half-year. Fifth, to capture potential effects on ratings attributable to retention decisions, the dummy

variable *Second HY* marks observations from the second half-year of the season after which players are either promoted or dismissed. Anecdotal evidence indicates that some coaches might have assigned more generous grades to prevent the dismissal of a player or to justify his retention.

Methods

Our predictions are tested in three steps. First, we search for general evidence of optimistic (H_1) and positivity (H_2) biases in the coaches' evaluations by conducting paired *t*-tests and Wilcoxon sign-rank tests comparing the means of *predicted* versus actual performance and of *remembered* versus actual performance. The results are graphically supported by kernel density plots. Second, we explore the bivariate relationships between our dependent variables, *Optimistic Bias* and *Positivity Bias*, and the independent variables by calculating correlation coefficients. We also analyze correlations among the regressors to identify potential multicollinearity problems. Third, in models 1 and 2 and models 3 and 4, respectively, we test the hypothesized moderators of the optimistic and positivity biases (H_3-H_5) by estimating four multivariate OLS regressions using the biases as dependent variables. In both cases, the first model excludes controls, which are added into the second model. We also compute robust standard errors (by clustering over players) to minimize potential bias from heteroscedasticity, and standardized regression coefficients (betas) to facilitate effect size comparisons among our independent variables.

Results

Tests of hypotheses

First, we examine the theoretically motivated existence of an optimistic bias (hypothesis 1) and positivity bias (hypothesis 2) in employee ratings using paired *t*-tests and Wilcoxon sign-rank tests. The results in Table 2 reinforce both propositions. First, as postulated by hypothesis 1, a player's predicted performance is on average rated more favorably than his

actual performance. The mean difference of +1.999 on the 10-point Likert scale is significant in both parametric and nonparametric tests across all age groups at $p < .001$ except for one case in which $z(11) = 2.934$, $p = .003$. Likewise, hypothesis 2 is supported by a positive mean difference between remembered and actual performance of +1.092, which meets the $p < .001$ criterion in all age groups except two in which $z(11) = 2.936$, $p = .003$ and $z(42) = 2.496$, $p = .013$. Moreover, the average ratings of predicted (8.164) and remembered performance (7.257) correspond to the 99th and 89th percentile of the ratings of actual performance.

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

As also shown in Figure 2, 95.1% (optimistic bias) and 88.8% (positivity bias) of the differences are greater than 0, although interestingly, the prospective optimistic bias is on average almost twice as pronounced as the retrospective positivity bias.

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

Bivariate Results

We next compute pairwise correlations to explore the relation between the optimistic and positivity biases and ratee/rater characteristics. The results in Table 3 support hypothesis 3, contradict hypothesis 4, and provide mixed evidence on hypothesis 5. First, as anticipated by hypothesis 3, ratee age is positively correlated with both the optimistic (+.16, $p = .002$) and the positivity (+.24, $p = .000$) bias. However, the hypothesis 4 prediction of a negative relation between both rating distortions and rater tenure is refuted by the positive coefficients for the correlations between rater tenure and the optimistic (+.28, $p = .000$) or positivity (+.25, $p = .000$) bias. On the other hand, the hypothesis 5 assumption of a positive connection between relationship closeness and the optimistic (+.08, $p = .144$) and positivity (+.16, $p = .002$) biases is confirmed for the latter and at least not rejected for the former. The multicollinearity between the independent variables also seems manageable: at most, the

significantly positive correlation of the temporal measures *Relationship Closeness* and *Second HY* might be of concern.

--- Insert Table 3 about here ---

Multivariate Results

We then estimate two sets of two multivariate OLS models (with and without controls) using either the optimistic (Table 4) or the positivity (Table 5) bias as the dependent variable. These estimations support the bivariate results for hypotheses 3 and 4, but shed further doubt on hypothesis 5. For hypothesis 3, the regression coefficients confirm a positive influence of ratee age on both the optimistic (+.089, $t(164) = 2.65$, $p = .009$) and positivity (+.098, $t(164) = 5.16$, $p = .000$) biases, which increases to +.117 ($t(164) = 3.48$, $p = .001$) and +.111 ($t(164) = 5.36$, $p = .000$), respectively, when controls are added into models 2 and 4. That is, if a player becomes one year older, expected and remembered performance will deviate by an additional +.117 and +.111 points, respectively, from actual performance. We can thus not reject hypothesis 3. As regards hypothesis 4, however, the OLS estimates replicate the bivariate results of a positive effect of rater tenure on the optimistic (+.017, $t(164) = 4.55$, $p = .000$) and positivity (+.011, $t(164) = 4.68$, $p = .000$) biases that remains robust to the introduction of controls, We must therefore reject hypothesis 4. The multivariate results further reinforce our concerns about hypothesis 5: although the pairwise correlations support the predicted positive (albeit insignificant) relation, the OLS coefficients capturing the influence of relationship closeness on the optimistic (-.009, t(164) = -.50, n.s.) and positivity (-.010, t(164) = -.94, n.s.) biases are negative. Thus, we cannot uphold hypothesis 5.

--- Insert Tables 4 and 5 about here ---

As the standardized betas show, the effects of ratee age and rater tenure on the two biases are of similar magnitude and thus larger than the effects of relationship closeness or any control variable. For instance, one standard deviation change of ratee age and rater tenure leads to

.218 and .314 standard deviation changes of the optimistic bias (.281 and .278 changes, respectively, of the positivity bias) in Models 2 and 4. In the latter case, this impact is twice as pronounced as any control variable. Multicollinearity is a minor concern: the variance inflation factors of the three moderators do not exceed 2.22, which is below the common threshold of 5 (e.g., Sheather, 2009) or 10 (e.g., Kutner, Nachtsheim, & Neter, 2004). While the presented results are clustered over players, a clustering over coaches yields similar coefficients, yet at lower significance levels, possibly due to the limited number of trainers.

Discussion

The above analyses seek evidence of optimistic (H_1) and positivity (H_2) biases in performance appraisals that are moderated by ratee and rater characteristics, such as amount of ratee information (H_3), rater experience (H_4), and the closeness of the rater/ratee relationship (H_5). When deriving hypotheses 1 and 2, we argued that laboratory-based proofs of self-relevant optimistic and positivity biases and their extension to others suggest a positive distortion of prospective and retrospective employee ratings. Both these propositions are supported by the empirical finding that *predicted* (mean difference +1.999, p = .000) and *remembered* (mean difference +1.092, p = .000) performance ratings are significantly higher than *actual* (instantly reported) performance ratings. Most notably, the "umbrella of optimistic biases" is nearly twice as pronounced as the "rose-colored view of the past", which is analogous to the finding in "rosy view" studies that the anticipation of personal life events is even more joyful than their recollection (e.g. Wirtz, Kruger, Scollon, & Diener, 2003). This asymmetry in looking forward and back deserves mention and has already intrigued other researchers. Mitchell et al. (1997) propose two explanations: First, the frequent discrepancy between affective anticipation and factual occurrence creates a residual post-event disappointment that persists despite the rosy view. Second, in light of the commonly

observable expectation inflation, slightly lower on-line and post-event evaluations could simply be an inevitable regression-to-the-mean effect.

In hypothesis 3, we postulate that ratee age will have a positive influence on both the optimistic and the positivity bias, an effect that is in fact reinforced by the bi- and multivariate results. In line with Regan et al. (1995) and Shepperd et al. (2002), older players seem to receive favorable treatment because coaches can, ceteris paribus, draw on a richer pool of specific information to rationalize overoptimistic judgments. This argument appears plausible given that other future-oriented biases such as the planning fallacy are also attributed to selective (ab)use of detailed personalized knowledge (Buehler, Griffin, & Ross, 1994).

We must reject hypothesis 4, however, because the empirical findings show a positive impact of rater experience on both biases, an outcome that is puzzling at first glance. One possible interpretation along the lines of hypothesis 3 is that long-tenured coaches can draw on a rich pool of examples from academy players who developed better than expected. Again, the selective recall of such specific information would render it easier to rectify an overoptimistic outlook. We are also unable to uphold hypothesis 5 on the positive effect of relationship closeness, whose direction is confirmed by the pairwise correlations but shown by the multivariate OLS coefficients to be negative in all four models. One possible explanation is that the global prospective and retrospective performance ratings, once assigned after six months, constitute a (usually positive) reference point from which coaches do not substantially deviate thereafter – an evaluation pattern that is supported by anecdotal evidence from members of the academy. Consequently, the gap between predicted, remembered, and actual performance declines in subsequent half-years.

Admittedly, alternative interpretations are possible. For example, the optimistic and positivity biases might simply mirror a different appraisal purpose for the continuously

assigned ratings of actual performance than for the summary ratings of predicted and remembered performance. Only the latter two coincide with promotion decisions, at least after the second half-year. In that case, however, the effect of the variable *Second HY* should not be insignificant, which it is. Moreover, because one coach grades all players in one age group, all assessments are relative, meaning there is no pressure to artificially inflate the global ratings, especially since there are no fixed "cut-off grades" that preclude a promotion. Thus, several arguments weaken the appraisal purpose explanation.

Another possibility is that the more official summary ratings of predicted and remembered performance are inflated because they reach a broader audience, meaning that a coach might positively manipulate his impressions to appear as a competent manager or at least to "avoid hanging dirty laundry out in public" (Longenecker et al., 1987, p. 189). Yet this interpretation is made less plausible by the existence of an Intranet portal on which academy staff, scouts, and coaches can transparently retrace all the assessments and statistics for every single player.

Practical Implications

The conclusions drawn from this analysis of coaches' evaluations of their young soccer players have implications for performance appraisals in a business context. In particular, our findings suggest that predicted and remembered performance ratings are a distorted reflection of actual employee performance. Companies might thus take steps to reduce the retrospective positivity of subordinate evaluations, such as introducing a system of continuous reporting and archiving of intermediate ratings upon which supervisors can later draw. Nevertheless, the mere establishment of such a system might not eliminate rose-colored views of past employee performance, meaning that they should also consider complementing it with forced distribution mechanisms or at least a sensitization of supervisors to the issue. For instance,

prior research shows that explicit training, advice, and discussions with raters can diminish psychometric errors such as halo or leniency effects (Bernardin, 1978; Smith, 1986).

In general, our results provide yet another example of optimistically biased prognoses in an organizational setting, which include discounts to costs and completion times of major initiatives (Lovallo & Kahneman, 2003), overestimations of investment returns (Malmendier & Tate, 2005), and excess entries into competitive markets (Camerer & Lovallo, 1999). Such unrealistic forecasts frequently lead to the continuation of economically unviable projects or, as here, to the promotion of human resources with limited potential. This problem is especially obvious when responsible supervisors periodically rotate so as not to incur the implicit costs of their previous optimistic decisions.

Limitations and Future Directions

Admittedly, despite positively contributing to the performance appraisal literature by providing evidence on the intertemporal nature of rating errors, this study is subject to several limitations, some of which highlight avenues for further investigation. First, the study context (soccer) differs from the setting for which its findings are primarily relevant (business). However, because the academy's appraisal and promotion system resembles typical company practices, we assume that the results from this controlled real-life laboratory of professional sports (Jung et al., 2012) are generalizable to the business setting. Nevertheless, we cannot eliminate the possibility that some effects may arise from soccer-specific particularities, suggesting a need to replicate our findings in a business environment.

Second, the sample is limited in size and features a comparably homogenous group of male, competitive, and predominantly German youngsters. Although such homogeneity minimizes issues related to unobservables, we cannot be certain that our insights apply equally across all age groups, countries, or genders, which can be a source of bias in personnel evaluations (see Eagly, Makhijani, & Klonsky, 1992, for gender). Future studies

might therefore increase the sample size and diversify the subject pool to verify the validity of our generalizations.

Third, our dataset is rather weakly defined longitudinally, so expanding the observation period to more than three years would facilitate detection of unrealistic optimism at the subject level by comparing individual expectations with actual long-term outcomes (see Armor & Taylor, 1998). On that basis, future studies could determine the stability of the investigated rating biases over time.

Conclusions

In this examination of the intertemporal variations in coaches' ratings of young elite soccer players, the subjects' *predicted* and *remembered* performance ratings significantly exceed their *actual* performance ratings. This effect is most pronounced when a considerable amount of information is available on the ratee and when raters are experienced. Moreover, the positive deviation of predicted from actual performance is even more pronounced than for remembered performance, which suggests an asymmetry in looking forward and backward.

Based on the theoretical argument that both prospective optimistic and retrospective positivity biases can extend to judgments on others, we link our empirical findings to cognitive distortions in autobiographical memory. We subsequently transpose the bi- and multivariate results to the business domain, suggesting that employee evaluations might be subject to systematic intertemporal bias that, like any rating error, would imply significant costs for the company (Heneman et al., 1987). We then recommend steps to mitigate these effects, including supervisor training or systems that facilitate the continuous reporting of intermediate grades.

In addition to offering practical suggestions, we contribute to the performance appraisal literature by advancing current understanding of the intertemporal nature of rating errors. We also provide real-world evidence of two cognitive biases that have to date been studied

predominantly in a laboratory setting. In doing so, we lay down a useful foundation for future research, which might consider testing our hypotheses in a business environment, enhancing the sample size and variety, and/or employing a longitudinal design to overcome any study limitations and strengthen the generalizability of our conclusions.

References

Antonioni, D., & Park, H. (2001). The Effects of Personality Similarity on Peer Ratings of Contextual Work Behaviors. *Personnel Psychology, 54*, 331–360.

Armor, D. A., & Taylor, S. E. (1998). Situated Optimism: Specific Outcome Expectancies and Self-Regulation. In M. P. Zanna & J. M. Olson (Eds.), *Advances in Experimental Social Psychology* (Vol. 30, pp. 309–379). London: Academic Press.

Aron, A. P., & Aron, E. N. (1986). *Love and the Expansion of Self: Understanding Attraction and Satisfaction*. New York: Hemisphere Publishing Corp/Harper & Row Publishers.

Aron, A. P., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close Relationships as Including Other in the Self. *Journal of Personality and Social Psychology, 60*, 241–253.

Aron, A. P., Lewandowski, G. W., Mashek, D. J., & Aron, E. N. (2013). The Self-Expansion Model of Motivation and Cognition in Close Relationships. In J. A. Simpson & L. Campbell (Eds.), *The Oxford Handbook of Close Relationships* (pp. 90–115). Oxford: Oxford University Press.

Aron, A. P., Mashek, D. J., & Aron, E. N. (2004). Closeness as Including Other in the Self. In D. J. Mashek & A. P. Aron (Eds.), *Handbook of Closeness and Intimacy* (pp. 27–42). Mahwah: Lawrence Erlbaum Associates.

Bahrick, H. P., Hall, L. K., & Berger, S. A. (1996). Accuracy and Distortion in Memory for High School Grades. *Psychological Science, 7*, 265–271.

Balzer, W. K., & Sulsky, L. M. (1992). Halo and Performance Appraisal Research: A Critical Examination. *Journal of Applied Psychology, 77*, 775–785.

Baumeister, R. F., & Newman, L. S. (1994). How Stories Make Sense of Personal Experiences: Motives that Shape Autobiographical Narratives. *Personality and Social Psychology Bulletin, 20*, 676–690.

Bernardin, H. J. (1978). Effects of Rater Training on Leniency and Halo Errors in Student Ratings of Instructors. *Journal of Applied Psychology, 63*, 301–308.

Bernardin, H. J., & Villanova, P. (1986). Performance Appraisal. In E. Locke (Ed.), *Generalizing from Laboratory to Field Settings* (pp. 43–62). Lexington: Lexington Books.

Berntsen, D., & Rubin, D. C. (2004). Cultural Life Scripts Structure Recall from Autobiographical Memory. *Memory & Cognition, 32*, 427–442.

Bol, J. C. (2011). The Determinants and Performance Effects of Managers' Performance Evaluation Biases. *The Accounting Review, 86*, 1549–1575.

Boucher, J., & Osgood, C. E. (1969). The Pollyanna Hypothesis. *Journal of Verbal Learning and Verbal Behavior, 8*, 1–8.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "Planning Fallacy": Why People Underestimate Their Task Completion Times. *Journal of Personality and Social Psychology, 67*, 366–381.

Burger, J. M., & Burns, L. (1988). The Illusion of Unique Invulnerability and the Use of Effective Contraception. *Personality and Social Psychology Bulletin, 14*, 264–270.

Burger, J. M., & Palmer, M. L. (1992). Changes in and Generalization of Unrealistic Optimism Following Experiences with Stressful Events: Reactions to the 1989 California Earthquake. *Personality and Social Psychology Bulletin, 18*, 39–43.

Camerer, C., & Lovallo, D. (1999). Overconfidence and Excess Entry: An Experimental Approach. *American Economic Review, 89*, 306–318.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in Social Comparative Judgments: The Role of Nonmotivated Factors in Above-Average and Comparative-Optimism Effects. *Psychological Bulletin, 130*, 813–838.

Christensen, M. K. (2009). "An Eye for Talent": Talent Identification and the "Practical Sense" of Top-Level Soccer Coaches. *Sociology of Sport Journal, 26*, 365–382.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple Uses of Performance Appraisal: Prevalence and Correlates. *Journal of Applied Psychology, 74*, 130–135.

Cooper, W. H. (1981). Ubiquitous Halo. *Psychological Bulletin, 90*, 218–244.

Crary, W. G. (1966). Reactions to Incongruent Self-Experiences. *Journal of Consulting Psychology, 30*, 246–252.

Croyle, R. T., Loftus, E. F., Barger, S. D., Sun, Y. C., Hart, M., & Gettig, J. (2006). How Well Do People Recall Risk Factor Test Results? Accuracy and Bias among Cholesterol Screening Participants. *Health Psychology, 25*, 425–432.

Deutsche Fußball Liga (2010). *10 Jahre Leistungszentren*. Frankfurt am Main: Deutsche Fußball Liga.

Deutscher Fußball-Bund (2010). *Richtlinien für die Errichtung und Unterhaltung von Leistungszentren der Teilnehmer der Lizenzligen* (Anhang V zur Lizensierungsordnung). Frankfurt am Main: Deutscher Fußball-Bund.

Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F. J., Bachl, N., & Pigozzi, F. (2007). Performance Characteristics According to Playing Position in Elite Soccer. *International Journal of Sports Medicine, 28*, 222–227.

Drake, R. A. (1984). Lateral Asymmetry of Personal Optimism. *Journal of Research in Personality, 18*, 497–507.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the Evaluation of Leaders: A Meta-Analysis. *Psychological Bulletin, 111*, 3–22.

Ghosh, S., & Waldman, M. (2010). Standard Promotion Practices versus Up-or-Out Contracts. *Rand Journal of Economics, 41*, 301–325.

Greenwald, A. G. (1980). The Totalitarian Ego: Fabrication and Revision of Personal History. *American Psychologist, 35*, 603–618.

Helsen, W. F., van Winckel, J., & Williams, A. M. (2005). The Relative Age Effect in Youth Soccer Across Europe. *Journal of Sports Sciences, 23*, 629–636.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do Moderators of the Optimistic Bias Affect Personal or Target Risk Estimates? A Review of the Literature. *Personality and Social Psychology Review, 5*, 74–95.

Heneman, R. L., Moore, M. L., & Wexley, K. N. (1987). Performance-Rating Accuracy: A Critical Review. *Journal of Business Research, 15*, 431–448.

Jacobs, R., & Kozlowski, S. W. (1985). A Closer Look at Halo Error in Performance Ratings. *Academy of Management Journal, 28*, 201–212.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and Reconstructing the Past and the Future in the Present. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of Motivation and Cognition: Foundations of Social Behavior* (Vol. 2, pp. 482–526). New York: Guilford Press.

Jung, V., Schmidt, S. L., & Torgler, B. (2012). *Antecedents of Attitudes Towards Risky Career Choices* (CREMA Working Paper No. 2012-20). Basel: CREMA.

Kahneman, D. (2003). Objective Happiness. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-Being: Foundations of Hedonic Psychology* (pp. 3–25). New York: Russell Sage Foundation.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of Experienced Utility. *Quarterly Journal of Economics, 112*, 375–406.

Konrath, S. H., & Ross, M. (2003). Our Glories, Our Shames: Expanding the Self in Temporal Self-Appraisal Theory. Paper presented at the 111th Annual Meeting of the American Psychological Society, August 7–10, Toronto, Ontario, Canada.

Kraiger, K., & Ford, J. K. (1985). A Meta-Analysis of Ratee Race Effects in Performance Ratings. *Journal of Applied Psychology, 70*, 56–65.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models* (4th ed.). New York: McGraw-Hill Irwin.

Landy, F. J., & Farr, J. L. (1980). Performance Rating. *Psychological Bulletin, 87*, 72–107.

Langer, E. J. (1975). The Illusion of Control. *Journal of Personality and Social Psychology, 32*, 311–328.

Latham, G. P., & Wexley, K. N. (1982). *Increasing Productivity Through Performance Appraisal*. Reading: Addison-Wesley.

Longenecker, C. O., Sims, H. P., & Gioia, D. A. (1987). Behind the Mask: The Politics of Employee Appraisal. *Academy of Management Executive, 1*, 183–193.

Lovallo, D., & Kahneman, D. (2003). Delusions of Success. How Optimism Undermines Executives' Decisions. *Harvard Business Review, 81*(7), 56–63.

Malmendier, U., & Tate, G. (2005). CEO Overconfidence and Corporate Investment. *Journal of Finance, 60*, 2661–2700.

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna Principle: Selectivity in Language, Memory, and Thought*. Cambridge: Schenkman.

McKenna, F. P. (1993). It Won't Happen to Me: Unrealistic Optimism or Illusion of Control? *British Journal of Psychology, 84*, 39–50.

McShane, S. L. (1990). Two Tests of Direct Gender Bias in Job Evaluation Ratings. *Journal of Occupational Psychology, 63*, 129–140.

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal Adjustments in the Evaluation of Events: The "Rosy View." *Journal of Experimental Social Psychology, 33*, 421–448.

Murdock Jr., B. B. (1962). The Serial Position Effect of Free Recall. *Journal of Experimental Psychology, 64*, 482–488.

Musch, J., & Grondin, S. (2001). Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport. *Developmental Review, 21*, 147–167.

Ng, K. Y., Koh, C., Ang, S., Kennedy, J. C., & Chan, K. Y. (2011). Rating Leniency and Halo in Multisource Feedback Ratings: Testing Cultural Assumptions of Power Distance and Individualism-Collectivism. *Journal of Applied Psychology, 96,* 1033–1044.

Perloff, L. S., & Fetzer, B. K. (1986). Self-Other Judgments and Perceived Vulnerability to Victimization. *Journal of Personality and Social Psychology, 50*, 502–510.

Reb, J., & Cropanzano, R. (2007). Evaluating Dynamic Performance: The Influence of Salient Gestalt Characteristics on Performance Ratings. *Journal of Applied Psychology, 92*, 490–499.

Regan, P. C., Snyder, M., & Kassin, S. M. (1995). Unrealistic Optimism: Self-Enhancement or Person Positivity? *Personality and Social Psychology Bulletin, 21*, 1073–1082.

Ross, M., & Wilson, A. E. (2002). It Feels Like Yesterday: Self-Esteem, Valence of Personal Past Experiences, and Judgments of Subjective Distance. *Journal of Personality and Social Psychology, 82*, 792–803.

Rothwell, W. J. (2010). *Effective Succession Planning: Ensuring Leadership Continuity and Building Talent from Within* (4th ed.). New York: Amacom.

Rubin, D. C., & Berntsen, D. (2003). Life Scripts Help to Maintain Autobiographical Memories of Highly Positive, but Not Highly Negative, Events. *Memory & Cognition, 31*, 1–14.

Sande, G. N., Goethals, G. R., & Radloff, C. E. (1988). Perceiving One's Own Traits and Others': The Multifaceted Self. *Journal of Personality and Social Psychology, 54*, 13–20.

Schuh, A. J. (1973). Effects of Interview Rating Form Content and Rater Experience on the Evaluation of a Job Applicant. *Personnel Psychology, 26*, 251–260.

Sears, D. O. (1983). The Person-Positivity Bias. *Journal of Personality and Social Psychology, 44*, 233–250.

Sedikides, C., & Green, J. D. (2009). Memory as a Self-Protective Mechanism. *Social and Personality Psychology Compass, 3*, 1055–1068.

Sedikides, C., & Gregg, A. P. (2007). Portraits of the Self. In M. A. Hogg & J. Cooper (Eds.), *The Sage Handbook of Social Psychology* (pp. 93–122). Thousand Oaks: Sage Publications.

Sedikides, C., & Strube, M. J. (1997). Self-Evaluation: To Thine Own Self Be Good, to Thine Own Self Be Sure, to Thine Own Self Be True, and to Thine Own Self Be Better. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 29, pp. 209–269). Bingley: Emerald Publishing.

Sheather, S. (2009). *A Modern Approach to Regression with R*. New York: Springer.

Shepperd, J. A., Carroll, P., Grace, J., & Terry, M. (2002). Exploring the Causes of Comparative Optimism. *Psychologica Belgica, 42*, 65–98.

Skowronski, J. J. (2011). The Positivity Bias and the Fading Affect Bias in Autobiographical Memory. A Self-Motives Perspective. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of Self-Enhancement and Self-Protection* (pp. 211–233). New York: Guilford Press.

Smith, D. E. (1986). Training Programs for Performance Appraisal: A Review. *Academy of Management Review, 11*, 22–40.

Spence, J. R., & Keeping, L. M. (2010). The Impact of Non-Performance Information on Ratings of Job Performance: A Policy-Capturing Approach. *Journal of Organizational Behavior, 31*, 587–608.

Stauffer, J. M., & Buckley, M. R. (2005). The Existence and Nature of Racial Bias in Supervisory Ratings. *Journal of Applied Psychology*, *90*, 586–591.

Steiner, D., & Rain, J. S. (1989). Immediate and Delayed Primacy and Recency Effects in Performance Evaluation. *Journal of Applied Psychology, 74*, 136–142.

Symons, C. S., & Johnson, B. T. (1997). The Self-Reference Effect in Memory: A Meta-Analysis. *Psychological Bulletin, 121*, 371–394.

Taylor, S. E. (1991). Asymmetrical Effects of Positive and Negative Events: The Mobilization-Minimization Hypothesis. *Psychological Bulletin, 110*, 67–85.

Taylor, S. E., & Brown, J. D. (1988). Illusion and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin, 103*, 193–210.

Tichy, G. (2004). The Over-Optimism Among Experts in Assessment and Foresight. *Technological Forecasting and Social Change, 71*, 341–363.

Tiger, L. (1979). *Optimism: The Biology of Hope*. New York: Simon & Schuster.

Van Hoecke, J., Schoukens, H., Simm, S., Isakowitz, M., & de Sutter, R. (2011). The Classification of Youth Academies in German Professional Football Clubs. In H. Gammelsaeter & G. Bielons (Eds.), *Proceedings of the 19th Conference of the*

European Association for Sport Management (pp. 339–340). Madrid: GB Creation & Advice Consulting.

Walker, W. R., & Skowronski, J. J. (2009). The Fading Affect Bias: But What the Hell Is It For? *Applied Cognitive Psychology, 23*, 1122–1136.

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is Pleasant – and Memory Helps to Keep It That Way! *Review of General Psychology, 7*, 203–210.

Walker, W. R., Vogl, R. J., & Thompson, C. P. (1997). Autobiographical Memory: Unpleasantness Fades Faster than Pleasantness over Time. *Applied Cognitive Psychology, 11*, 399–413.

Weinstein, N. D. (1980). Unrealistic Optimism About Future Life Events. *Journal of Personality & Social Psychology, 39*, 806–820.

Weinstein, N. D. (1989). Perceptions of Personal Susceptibility to Harm. In V. M. Hays, G. W. Albee, & S. F. Schneider (Eds.), *Primary Prevention of AIDS: Psychological Approaches* (pp. 142–167). Newbury Park: Sage.

Weinstein, N. D., & Lachendro, E. (1982). Egocentrism as a Source of Unrealistic Optimism. *Personality and Social Psychology Bulletin, 8*, 195–200.

White, H., Coppola, H. A., & Multunas, N. K. (2008). Remembering Others: Using Life Scripts to Access Positive but Not Negative Information. *American Journal of Psychology, 121*, 293–318.

Wilson, A. E., & Ross, M. (2000). The Frequency of Temporal-Self and Social Comparisons in People's Personal Appraisals. *Journal of Personality and Social Psychology, 78*, 928–942.

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to Do on Spring Break? The Role of Predicted, On-Line, and Remembered Experience in Future Choice. *Psychological Science, 14*, 520–524.

Woodfield, R. L., & Viney, L. L. (1984). A Personal Construct Approach to the Conjugally Bereaved Woman. *Omega-Journal of Death and Dying, 15*, 1–13.

Zaleskiewicz, T. (2011). Financial Forecasts During the Crisis: Were Experts More Accurate than Laypeople? *Journal of Economic Psychology, 32*, 384–390.

Table 1

Descriptive Summary Statistics of Variables

Variable	*Mean*	*Median*	*SD*	*Min*	*Max*
Dependent variables					
Optimistic Bias	1.999	2.000	1.171	-2.857	4.833
Positivity Bias	1.092	1.100	.860	-1.833	3.833
Independent Variables					
Ratee Age	14.188	13.611	2.186	10.197	19.210
Rater Tenure	50.301	51.045	20.170	6.013	83.980
Relationship Closeness	9.156	6.000	4.943	6.000	36.000
Control Variables					
Goalkeeper[a]	.063	0	.243	0	1
Defender[a]	.293	0	.456	0	1
Striker[a]	.205	0	.405	0	1
Flexible Position[a]	.066	0	.248	0	1
Relative Age	134.764	112	94.244	3	362
Foreign[a]	.260	0	.439	0	1
Last Game	-.075	-.091	1.252	-4.444	3.4
Second HY[a]	.405	0	.492	0	1

[a]Dummy variable (1 = goalkeeper/ defender/ striker/ flexible position/ foreigner/ observation from 2nd half-year).

Table 2

Hypothesis Tests of Mean Differences Between Predicted, Remembered, and Actual Performance Ratings by Age Group

Age Group	Obser-vations	Predicted Performance		Remembered Performance		Actual Performance		H_0: Optimistic Bias = 0 (predicted-actual performance)				H_0: Positivity Bias = 0 (remembered-actual performance)			
		M	*SE*	*M*	*SE*	*M*	*SE*	*t*-test[a]		Wilcoxon test[b]		*t*-test[a]		Wilcoxon test[b]	
								t-value	*p*-value	*z*-score	*p*-value	*t*-value	*p*-value	*z*-score	*p*-value
< 12	42	7.929	.208	7.000	.132	6.703	.102	5.767	.000	4.390	.000	2.311	.013	2.496	.013
12	85	7.894	.166	7.047	.103	5.890	.069	13.347	.000	7.314	.000	13.386	.000	7.673	.000
13	78	8.590	.084	7.372	.082	6.325	.095	20.969	.000	7.661	.000	12.963	.000	7.442	.000
14	69	8.551	.078	7.971	.080	6.782	.089	18.144	.000	7.220	.000	14.493	.000	7.074	.000
15	11	8.545	.282	8.091	.211	6.273	.368	6.511	.000	2.934	.003	7.615	.000	2.936	.003
16	18	7.833	.326	6.667	.313	5.122	.147	9.336	.000	3.681	.000	6.136	.000	3.509	.000
17	22	8.273	.164	6.864	.178	5.619	.140	12.811	.000	4.107	.000	5.812	.000	3.685	.000
> 17	40	7.475	.143	6.775	.131	5.554	.105	13.190	.000	5.498	.000	8.578	.000	5.231	.000
Σ	365	8.164	.060	7.257	.049	6.166	.046	32.598	.000	16.001	.000	24.249	.000	15.322	.000

Note. Means and standard errors for predicted, remembered, and actual performance are measured on 10-point Likert scale.
[a]*t*-test of paired differences. [b]Wilcoxon sign-rank test of paired differences.

Table 3

Correlations Between Variables

		1	2	3	4	5	6	7	8	9	10	11	12
Dependent Variables													
1	*Optimistic Bias*												
2	*Positivity Bias*	**.64**											
Independent Variables													
3	*Ratee Age*	**.16**	**.24**										
4	*Rater Tenure*	**.28**	**.25**	-.04									
5	*Relationship Closeness*	.08	**.16**	**.33**	**.33**								
Control Variables													
6	*Goalkeeper*[a]	-.03	-.10	.04	.02	.03							
7	*Defender*[a]	**-.12**	-.03	.08	.03	.03	**-.17**						
8	*Striker*[a]	-.04	-.01	.01	-.08	-.04	**-.13**	**-.33**					
9	*Flexible Position*[a]	.03	-.06	**-.27**	-.01	-.06	-.07	**-.17**	**-.13**				
10	*Relative Age*	.07	.05	**-.20**	-.03	-.03	**-.10**	.08	.06	-.05			
11	*Foreign*[a]	.06	.05	**.11**	-.05	-.05	-.08	**-.19**	**.13**	-.01	.05		
12	*Last Game*	.06	.07	**.13**	-.03	**.12**	-.03	.00	.01	-.06	.05	-.06	
13	*Second HY*[a]	.02	**.11**	**.13**	**.15**	**.58**	.02	-.02	.04	-.04	-.00	.03	.07

Note. Bold correlations are significant at the $p < .05$ level.
[a]Dummy variable (1 = goalkeeper/ defender/ striker/ flexible position/ foreigner/ observation from 2nd half-year).

Table 4

Effects of Ratee and Rater Characteristics on the Optimistic Bias

Variable	Model 1			Model 2		
	b	Beta	VIF	b	Beta	VIF
Independent variables						
Ratee Age	.089**	.165	1.23	.117**	.219	1.50
	(.033)			(.034)		
Rater Tenure	.017***	.294	1.31	.018***	.314	1.34
	(.004)			(.004)		
Relationship Closeness	-.009	-.037	2.14	-.015	-.065	2.22
	(.017)			(.019)		
Control variables						
Goalkeeper[a]				-.351[†]	-.073	1.12
				(.182)		
Defender[a]				-.489**	-.190	1.34
				(.144)		
Striker[a]				-.260	-.090	1.27
				(.159)		
Flexible Position[a]				.158	.034	1.25
				(.268)		
Relative Age				.002**	.124	1.11
				(.001)		
Foreign[a]				.029	.011	1.10
				(.134)		
Last Game				.044	.047	1.07
				(.046)		
Second HY[a]				.133	.056	3.52
				(.209)		
Constant	-.137			-.509		
Adjusted R^2	.136			.185		
N	365			365		

Note. b = unstandardized regression coefficient with robust standard error in parentheses. Beta = standardized regression coefficient. VIF = variance inflation factor. Standard errors are adjusted for clustering of observations by subject.

[a]Dummy variable (1=goalkeeper/ defender/ striker/ flexible position/ foreigner/ observation from 2nd half-year).

† $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5

Effects of Ratee and Rater Characteristics on the Positivity Bias

Variable	Model 3			Model 4		
	b	Beta	VIF	b	Beta	VIF
Independent Variables						
Ratee Age	.098***	.248	1.23	.111***	.281	1.50
	(.019)			(.021)		
Rater Tenure	.011***	.267	1.31	.012***	.278	1.34
	(.002)			(.002)		
Relationship Closeness	-.010	-.060	2.14	-.013	-.075	2.22
	(.011)			(.012)		
Control Variables						
Goalkeeper[a]				-.461*	-.130	1.12
				(.199)		
Defender[a]				-.211*	-.112	1.34
				(.104)		
Striker[a]				-.123	.058	1.27
				(.119)		
Flexible Position[a]				-.028	-.008	1.25
				(.180)		
Relative Age				.001*	.111	1.11
				(5.0e-4)		
Foreign[a]				.007	.004	1.10
				(.100)		
Last Game				.027	.040	1.07
				(.031)		
Second HY[a]				.067	.038	3.52
				(.145)		
Constant	-.707			-.902		
Adjusted R^2	.136			.171		
N	365			365		

Note. b = unstandardized regression coefficient with robust standard error in parentheses. Beta = standardized regression coefficient. VIF = variance inflation factor. Standard errors are adjusted for clustering of observations by subject.

[a]Dummy variable (1=goalkeeper/ defender/ striker/ flexible position/ foreigner/ observation from 2nd half-year).

† $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$.

	Development until 0.5 years later		Development until 1 year later		Development until 1.5 years later	
	t_0	$t_{+0.5}$	t_0	t_{+1}	t_0	$t_{+1.5}$
Actual Performance Ratings (*M* and 95% confidence interval)	6.43 **6.32** 6.21	6.47 **6.35** 6.23	6.74 **6.60** 6.47	6.68 **6.49** 6.30	6.71 **6.54** 6.37	6.54 **6.33** 6.12
Observations	202		107		79	
p ($t_{+x} > t_0$)	.328		.866		.940	

Figure 1. Development of actual performance ratings on 10-Point Likert scale over time. *p* ($t_{+x} > t_0$) is assessed with *t*-test of paired differences. Number of observations and actual performance in t_0 vary because sample becomes successively smaller as development intervals become longer.

Optimistic Bias							
%	1.6	3.3	13.7	31.8	31.5	15.1	3.0
cumulative %	1.6	4.9	18.6	50.4	81.9	97.0	100.0
Positivity Bias							
%	1.4	9.9	37.3	37.5	12.6	1.4	0.0
cumulative %	1.4	11.2	48.5	86.0	98.6	100.0	100.0

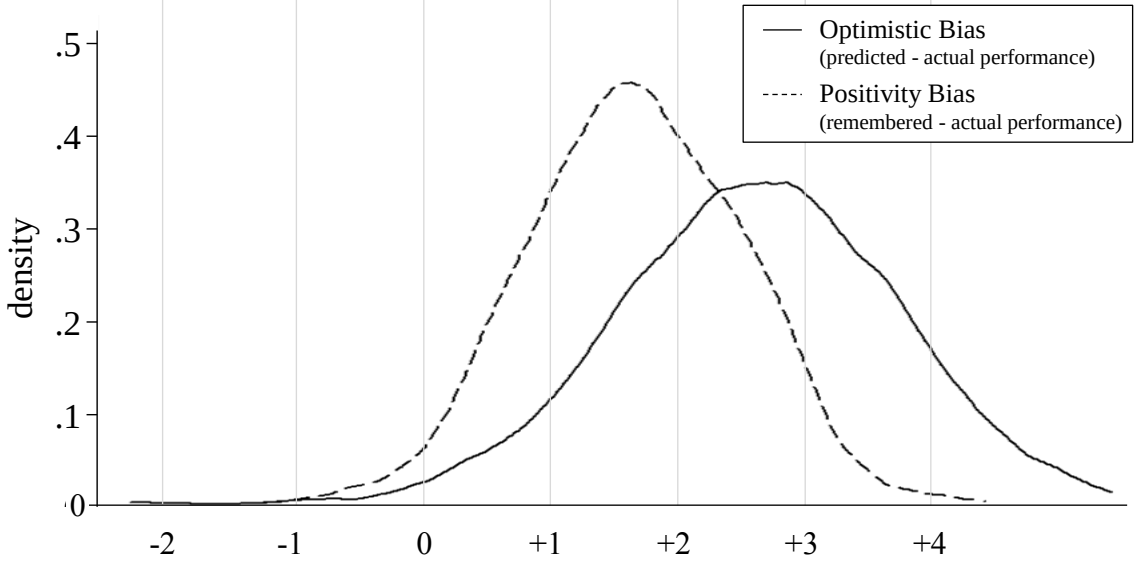

Figure 2. Distribution of the dependent variables (density kernel: Epanechnikov).