Transparency and biases in subjective performance evaluation

Jan Bouwens[1], **Christian Hofmann**[2], **Christopher Lechner**[2*]

Abstract: One of the major functions of accounting is to report unbiased numbers. In this paper, we demonstrate how favoritism may bias results and also how greater transparency can mitigate this bias. We examine whether one subtle feature of transparency – the physical presence of a stakeholder during the evaluation process – mitigates favoritism biases in subjective evaluations. Using archival data from professional ski jumping, we find that, controlling for objective performance indicators, subjective evaluations suffer from favoritism. In particular, evaluators favor athletes of their own nationality and athletes that have a compatriot on the evaluation panel. We test our transparency hypothesis taking advantage of the situation where, during lockdowns, sports competitions took place without an audience on-site. We predict and provide evidence that the physical presence of an audience during the decision-making process is associated with lower levels of favoritism in subjective evaluations. As such, we contribute to the accounting literature by highlighting how transparency may decrease the likelihood that subjective evaluations are biased by favoritism.

Keywords: Subjective Performance Evaluation; Bias; Favoritism; Transparency; Attendance

JEL: D91; M40; M51; L83; Z20

Current version: July 2022

Acknowledgements: We thank Theresa Libby and Patrick Kampkötter for valuable comments. Christian Hofmann gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft – Project-ID 403041268 – TRR266.

Corresponding author, Mail: jbouwens@uva.nl

[1] University of Amsterdam, Amsterdam Business School Accounting section, Plantage Muidergracht 121018 TV Amsterdam, The Netherlands

[2] LMU Munich, Munich School of Management, Geschwister-Scholl-Platz 1, 80539 Munich, Germany.

1 Introduction

Performance measures proxy for the contribution of employees to firm value. In order to increase performance measure precision and sensitivity to effort firms complete their measurement system with subjective evaluation of observed performance (Baker et al., 1994; Murphy & Oyer, 2003; Gibbs et al., 2004). While such interpretation may enhance the precision and sensitivity, the human involvement of subjective evaluation may also increase noise and/or force the measure into insensitivity or oversensitivity (Bol, 2011). Relying on biased evaluations increases the cost of human capital and impairs efficient decision-making (Moers, 2005; Grabner et al., 2020). Therefore, we study how the structure of the performance evaluation can mitigate the likelihood of intentional or nonintentional biases to occur. More precisely, we examine how one feature of transparency during the evaluation process – the physical presence of a stakeholder – mitigates favoritism biases in subjective evaluations in professional sports.

A variety of empirical challenges hinders analyzing subjectivity in traditional settings. First, researchers usually cannot access internal subjective evaluations of employees' performance. Second, even if subjective evaluations are accessible, data on in-group membership to identify favoritism is hardly available, protected by privacy policies, or lacks diversity with respect to the in-group membership. Moreover, researchers have limited possibility to control for an objective performance and for exogenous factors that influence employees' performance. Third, the number of observations in those field studies is usually small and employees' tasks different and changing.

To overcome these challenges, we take advantage of the institutional characteristics of ski jumping. The setting of our study follows the quest of Bol (2008, p. 28): "Data limitations have traditionally been an obstacle as subjective elements are not easily studied using traditional datasets. However, considering how important an understanding of the role of subjectivity in compensation contracting is, management control research must strive to find less traditional ways to tackle these issues".[1] In

[1] Similarly, Gibbs et al. (2004, p. 434) state: "However, many interesting issues arising with subjectivity are not easily studied using traditional datasets. By its nature, subjectivity involves concepts that are qualitative and, of course, subjective. It raises behavioral issues, such as trust, and conflicts in perceptions between the employee and supervisor. Such issues require a different approach, involving either field research or survey methods. While the data collection methods that we have employed are difficult, we expect that further progress on understanding the complexities of incentive systems, including the use of subjectivity, can be made by collecting new types of variables to examine theoretical concepts that have to date been given limited study."

professional ski jumping, the subjective evaluations of the athlete's performance by five different evaluators are publicly available, including respective names and nationalities. In addition to the subjective evaluations, there exists an objective performance score (length of jump) and data on the exogenous conditions that influence the respective performances. Thus, we can control for objective performance and exogenous factors to examine if there exists a favoritism bias in the subjective evaluation. Finally, in contrast to work situations, the tasks stay constant, and the judges are independent in the sense that they do not have to work together with the individuals they evaluate in the future. Hence, evaluators in our setting are less influenced by reciprocity concerns arising from future interactions between them and the athletes.

We make use of two features of this setting. First, we exploit the Covid-19 pandemic induced natural experiment that sports competitions took place without the physical presence of an audience. Second, five judges from five different countries subjectively evaluate athletes' performance in ski jumping. Thus, judges regularly evaluate an athlete with the same nationality, while at the same time the remaining four judges evaluate an athlete who may have a compatriot on the panel. Conceptually, it allows us to investigate if the presence of a stakeholder moderates favoritism by interacting the physical presence of an audience with variables that capture favoritism in subjective evaluations.

We collect subjective evaluations of athletes' performances from international judges in professional ski jumping between the years 2017 and 2022. In total, our sample consists of 44,905 evaluations of 151 athletes from 158 judges. We argue that judges' evaluations are prone to two types of favoritism: (1) Direct favoritism bias, occurring when the judge evaluates an athlete, who comes from the same nation; (2) Indirect favoritism, occurring when the judge evaluates an athlete whose compatriot is next to him on the judging panel.

We find that, controlling for objective performance indicators, judges' evaluations are on average significantly higher, when the athlete and the judge are from the same nation. In addition, the other judges tend to give athlethes higher subjective evaluations too in the presence of a compatriot of the athlete on the panel. These findings suggest that judges tend to favor their fellow citizens in their evaluation despite the various mechanisms in place. Since the coefficients are rather small in absolute terms, our results suggest that the judges favor their compatriots to an extent that does not undermine their credibility to the extend that it would jeopardize their chances of getting selected as a jury member in the future.

We subsequently test and find transparency reduces favoritism causing biases in subjective evaluations. Based on social facilitation theory, we argue that the physical presence of a stakeholder raises the evaluator's arousal, which increases the likelihood for his dominant response (Cottrell et al., 1968), i.e., to provide an unbiased evaluation. Thus, evaluators face higher cost of considering their inherent preference to favor members of the evaluator's in-group. We find that the presence of an audience is associated with lower favoritism biases in the subjective evaluation of the athlete's style in ski jumping. Finally we find that the presence of an audience is positively associated with higher subjective evaluations per se. This result would suggest that evaluators are reluctant to under value performance in the face of audience.

Methodologically, our study follows those of Zitzewitz (2006) and Sandberg (2018).[2] Zitzewitz (2006) analyzes favoritism in subjective evaluations in figure skating and ski jumping, Sandberg (2018) investigates favoritism based on nationality and gender in dressage competitions. While both studies provide strong evidence that judges favor their fellow citizens in subjective evaluations, Sandberg (2018) does not find favoritism on the basis of gender.

We contribute to the accounting literature on subjective performance evaluation (Gibbs et al, 2004; Bol, 2008, 2011). Systematic biases in subjective evaluation can have detrimental effects for employees, e.g., lower motivation and higher turnover (Moers, 2005). Thus, establishing processes and practices that facilitate fair evaluations is a core topic of management accounting. A large body of work identifies biases in subjective evaluations. For instance, evaluators systematically inflate the agent's performance (Moers, 2005), provide too condensed evaluations of performance (Bol, 2011), or favor particular employees (Prendergast & Topel, 1996). We extend these prior studies by showing that favoritism bias not only emerges through evaluator's favoring of his in-group members, but also through the composition of the decision-making body. In

[2] Two recent studies also use data from ski jumping to examine question of economic interest. Krumer et al. (2021) replicate the study of Zitzewitz (2006) for a larger data sample ranging from 2010 – 2017 and find strong support for the favoritism of fellow citizen. More interestingly, the authors find a statistically significant negative association between this favoritism bias and the corruption perception index of their countries, indicating that favoritism is more pronounced among judges from more corrupt countries. Scholten et al. (2020) examine the effect of the evaluator's age on favoritism bias. They provide evidence that older judges favor athletes from the same nationality more than younger judges do.

addition, we provide evidence on moderators of biases in the subjective evaluation. Fehrenbacher et al. (2018) experimentally analyze if the decision mode moderates the spill-over bias in subjective performance evaluation. They differentiate between deliberate and intuitive decision mode and provide evidence that deliberate decision mode is associated with a lower spill-over bias. Complementing this finding, we show that favoritism biases in subjective evaluations can be mitigated through the physical presence of a stakeholder. In particular, we exploit the variation of a feature of transparency, namely whether the evaluation-process takes place in the physical presence of a stakeholder. Hence, as in Fehrenbacher et al. (2018), our emphasis is on the debiasing in subjective evaluations that should ultimately lead to fairer and better decisions.

Also, we contribute to the literature on the effects of social factors (Akerlof, 1980; Bernheim, 1994; Becker & Murphy, 2000). Prior research shows that a crowd can influence a decision-maker in accordance with their preferences (e.g., Garicano et al., 2005; Bryson et al., 2021). We complement this research by highlighting a benefit of having a crowd physically present, as this reduces inherent preferences of decision makers. Thus, depending on the type of task and competitive setting, effects of a crowd on subjective evaluations vary.

The study also contributes to the accounting literature for it is not clear in advance whether transparency enhances or deteriorates information. Gradwohl and Feddersen (2018) find that transparency hampers decision making because decision makers become wary to express their thoughts in front of an audience that may criticize these thoughts in the first place. Their reluctance to express their thoughts puts an impediment on the exchange of ideas, which, in turn, affects the creation of information. This reduction in the quality of the information subsequently affects the quality of the decision that is based on this information. In our study we find, on the contrary, that transparency improves decision making information as the presence of an audience mitigates bias that results from favoritism.

Finally, our study fits into the growing literature that uses sports settings to investigate accounting related questions (Eyring et al., 2021; Black & Vance, 2021, Ferguson, 2021). While, exploiting sports data has a long tradition in economics (e.g., Chapman & Southwick, 1991; Price & Wolfers, 2010; Parsons et al. 2011), also accounting researcher have taken advantage of those data more recently. For instance, Eyring et al. (2021) provide experimental evidence that reporting only relative performance information to football players dominates the provision of absolute or absolute and relative performance information in increasing their performance. Black and Vance (2021) show that managers overly rely on their initial evaluations for promotion decisions in professional baseball after controlling for performance signals. Finally, Ferguson (2021) shows that supervisors put too much weight on outcome metrics relative to its information content. Using data from Australian-rules football his findings suggest that supervisors are unable to appropriately distinguish between factors of luck and effort when subjectively evaluating individuals' performance.

The remainder of the paper proceeds as follows. Section 2 provides a review of the literature and develops the hypothesis. Section 3 introduces the setting. Section 4 elaborates on the research design. Section 5 presents the main results and additional analyses. Section 6 discusses implications of our study. Section 7 concludes.

2 Literature review and hypothesis development

Subjective performance evaluation

The subjective evaluation of an agent's performance by an individual or a group of individuals is an integral part in diverse organizational, educational, governmental, legal, and sports settings. In many organizations, the associated subjective performance measures – along with objective performance measures – form the basis for compensation, budget allocations, or investment decisions (Prendergast, 1999). Prior research in economics and accounting suggests that subjective measures are beneficial to overcome shortcomings of objective measures. For instance, the subjective evaluation allows for a more holistic performance assessment, mitigates incentive distortion, or reduces risk by filtering out uncontrollable factors (e.g., Gibbs et al., 2004; Prendergast, 1999). However, the subjective evaluation is – by its very nature – discretionary and thus not (or just hardly) enforceable by a third party, which introduces the potential of untruthful or distorted performance measurement. Due to the lack of enforceability, the evaluator may also renege on the performance criteria, thus increasing the risk of the performance evaluation for the agent (Bol, 2008).

Biases in subjective performance evaluation

Conceptually, the subjective evaluation reflects the evaluator's assessment of the agent's performance subject to error, where the error can be systematic or unsystematic.

Prior work studies the systematic error in subjective evaluation, commonly referred to as bias. For instance, the evaluator may overstate the agent's performance (leniency bias, Moers, 2005), provide compressed ratings that fail to differentiate between actual performance levels (centrality bias, Bol, 2011), be influenced by unrelated performance measures (spill-over bias, Bol & Smith, 2011), or consider personal preferences (favoritism bias, Prendergast & Topel, 1996). Our focus is on favoritism bias.

Experimental research in social psychology suggests that individuals favor members of their in-group over members of other groups, where the in-group is defined as the social group to which the individual psychologically identifies as being a member (Billig & Tajfel, 1973). Empirical evidence based on archival data proposes that favoritism is present in various settings (Fisman et al., 2017; Parson et al., 2011; Shayo & Zussman, 2011). For example, Du et al. (2012) provide evidence of favoritism in the annual performance evaluation of state-owned enterprises in China.

Intrinsic motivation and reputational considerations make favoritism costly to the evaluator. Fehr et al. (1997) suggest that fairness considerations influence behavior in moral hazard settings in the sense that a guilt-averse evaluator incurs a cost when providing biased evaluations. Individuals care about their reputation as the reputation affects internal and external labor market opportunities (e.g., Yermack, 2004; Levit & Malenko, 2016). Hence, providing biased evaluations reduces the evaluator's economic prospects as evaluators who provide biased evaluations might not be considered for future promotions. Indeed, Grabner et al. (2020) find that less condensed and thus more informative performance ratings are associated with a higher likelihood of promotion.

Biased performance evaluations are costly to organizations because they reduce the organization's productivity and increase the cost of human capital (Moers, 2005). When agents consider the evaluations to be implausible or unfair, their motivation suffers and turnover rises. Biased evaluations may also trigger costly influencing activities by agents in their desire to appear in a favorable light (Du et al., 2012). More generally, providing biased evaluations to decision makers adversely affects the efficient allocation of employees and resources to projects and tasks (Grabner et al., 2020).

To mitigate the negative consequences caused by biased evaluations, organizations resort to diverse accountability and transparency measures, including the use of objective performance measures, multiple evaluators, re-evaluations of the subjective evaluations and adjustments of the evaluator's own performance evaluation.

Prendergast and Topel (1996) study favoritism in an agency setting and analyze the effects of favoritism on compensation and organizational structure. They show that organizations respond to costly favoritism by deemphasizing incentive pay and resorting to inefficient bureaucratic rules when aggregating performance information. Consistently, Demere et al. (2019) show that calibration committees adjust evaluators' biased evaluations. In a field study, Grabner et al. (2020) find that evaluators' opportunistic upward biasing of subjective performance evaluations is mitigated through receiving lower performance assessments for themselves.

Consistent with the adage that "sunlight is the best disinfectant," transparency is often thought to reduce favoritism in subjective evaluations. Transparency relates to the feature that, ex post, an individual's decisions are revealed to third parties. Levy (2007) establishes that when the individual's decisions are not publicly available, the individual complies with preexisting biases. A more subtle feature of transparency is whether the evaluation-process takes place under the physical presence of a stakeholder. For example, organizations differ whether prospective auditors pitch their auditing services either to a broad group including the members of the audit committee and senior managers of the accounting department or only to the members of the audit committee. While in each case the audit committee – at least formally – chooses the new auditor, the auditors pitching to a broader group implies that the committee's questions are under scrutiny by the senior managers, potentially affecting the audit committee's questions and ultimate their (subjective) choice. Further examples are innovation challenges where employees pitch their ideas either in town-hall meetings in front of the workforce or only to the jury members; job placement, promotion, or tenure cases where candidates present themselves either in front of the whole department or just in front of the relevant decision makers; congressional hearings that take place either publicly with the press being present or behind closed doors.[3]

[3] The effect of a third party on the evaluator's favoritism bias is distinct from recent work that investigates how social pressure influences evaluator's subjective decisions, i.e., that crowds can pressure the evaluator's decisions in their favor (e.g., Garicano et al., 2005; Bryson et al., 2021; Scoppa, 2021).

Social facilitation theory

Social facilitation theory provides a reasoning for why the physical presence of others may affect the evaluators' performance assessment.[4] Triplett (1898), Travis (1925), Pessin (1933), and Pessin and Husband (1933), among others, find evidence for an audience effect that is, the presence of other individuals affects an individual's task performance. Zajonc (1965) suggests that the mere presence of an audience increases the individual's arousal and thus facilitates the emission of well-learned responses but impairs the acquisition of new responses. In more detail, Mullen et al. (1997) suggest that the effect of the presence of others on an individual's arousal varies with the type of situation and the type of others.[5] Their meta-analysis shows that independent of the type of situation, the presence of others significantly increases an individuals' arousal, measured by electrodermal responses.[6] Consistently, Cottrell et al. (1968) find that the presence of an audience (i.e., spectators) enhances the emission of dominant responses for trained individuals.[7] That is, in the face or direct observation by others affect how the individual performs a task.

Hypothesis development

The prior discussion highlights that the evaluator has a preference for in-group members and derives utility from favoring them in their subjective evaluations. However, favoring certain agents is costly to the evaluator when the evaluator has a sense of fairness and acts against their own moral standard and when being known as partial and unfair reduces internal and external labor market prospects. Then, the optimal bias in the evaluator's subjective performance evaluation is such that the marginal benefit from the bias equals the marginal cost.

---

[4] Conceptually, Akerlof (1980), Bernheim (1994), and Becker and Murphy (2000), among others, argue that social factors influence decision-making. Among those social factors, social facilitation theory refers to the presence of other individuals.

[5] Mullen et al. (1997) differentiate (1) between neutral and aversive settings and (2) between mere presence others, audience others, and coactor others, where the others differ whether they are engaged in the same behavior as the individual and present to monitor the individual's behavior. Mullen et al. (1997) predict that audience others increase arousal for both neutral and aversive settings. Our empirical setting is likely characterized by a neutral setting where audience others stimulate social monitoring.

[6] Social facilitation has been shown across species different from men (e.g., birds, ants, cockroaches, monkeys; see Ogura & Matsushima, 2011), which suggests that this phenomenon has an ancient common phylogenetically cause.

[7] Aiello and Douthitt (2001) discuss various theories that explain the mechanism underlying the social facilitation effect. Besides the arousal explanation for social facilitation, the presence of an audience may also trigger concerns about comparisons with others or a shift in cognitive processing capacity.

Following social facilitation theory, the physical presence of a third party raises the evaluator's arousal, particularly for a well-trained evaluator. More specifically, the presence of a stakeholder increases the salience of unbiased evaluations, thereby increasing the evaluator's cost from their preference to favor members of the evaluator's in-group. Consequently, the physical presence of a stakeholder increases the emission of dominant responses, which corresponds to providing less biased subjective performance evaluations.[8] In other words, in the direct presence of a stakeholder the evaluator is reluctant to make a decision that would be construed as favoritism. We summarize our expectation in the following hypothesis:

H1: The physical presence of a stakeholder is negatively associated with favoritism bias in subjective performance evaluation.

3 The setting

We investigate judges' subjective performance evaluations for professional athletes competing in ski jumping. Ski jumping is an Olympic sport since the very first Winter Olympics in 1924, mainly popular in Europe as well as in Japan, and governed by the international ski federation (FIS).

Athletes simultaneously optimize two different maximization problems, which are to fly as far as possible and with the best possible style. Thus, ski jumping is a two-task setting, where the athletes' performance on the first task (fly as far as possible) is evaluated objectively, while their performance on the second task (with the best possible style) is evaluated subjectively. Athletes perform their jumps on specially build ski jumping hills. The FIS distinguishes between three different hill sizes: normal hills, large hills and flying hills. Depending on the hill size, athletes can jump distances of more than 200 meters on flying hills.

A competition usually runs as follows: After a qualification, 50 athletes compete in the first round. The best performing 30 athletes continue to compete in the second round. The athlete with the highest aggregated score of the two rounds wins the competition. The overall score for each jump is calculated as the sum of the objective

[8] Our line of argumentation corresponds to Prendergast and Topel (1996). In their model, the management monitors the evaluator's assessment and penalizes deviations from the management's own assessment, whereas we argue that the physical presence of a third party increases the emotional cost of biasing for the evaluator.

score, the subjective score and the compensation for wind and gate conditions. The objective score is calculated as 60 points for reaching the K-point (construction point) plus the product of a 'meter value' and the difference between the distance and the K-point.

The FIS regulates the payout structure. The organizer of a men's World Cup must pay out at least 71,800 Swiss Francs in prize money for an individual competition of which 10,000 Swiss Francs go to the winner and 8,000 Swiss Francs to the second placed.

To facilitate fairness within a competition, jumpers are compensated (or punished) for unfavorable (favorable) exogenous conditions. The exogenous conditions are wind (strength and direction) and the gate from which the jumper starts. If the wind conditions are favorable, i.e., head wind, points are subtracted, if it is detrimental, i.e., tail wind, points are added proportional to the strength of the wind.

Five judges (in World Cup events: five judges of five different countries, one from the hosting country) subjectively evaluate the style of the jump. Thus, there are regularly situations, in which a judge evaluates an athlete from his own country while at the same time the remaining four judges evaluate an athlete who has a compatriot on the panel. The FIS (2018, p. 62) prescribes that they rate the "outer appearance of the succession of the jumpers' movements, from the end of the take off to the passing of the 'fall line' in the outrun, from the aspect of precision (timing), perfection (carrying out of the movements), stability (flight-position, outrun) and general impression." Each judge can give up to 20 points (in 0.5 increments), guidelines exist, but there is room for discretion. In addition, the FIS describes (FIS, 2018, p. 40): "The Jumping Judge shall follow the fundamental principle of scoring all jumps objectively as an expert. By appointing individuals as Jumping Judges, the FIS indicates its confidence that these Judges will follow rules of conduct to the best of their ability." Judges have to submit their scores into a mobile device within a few seconds after the jump. Communication among judges is prohibited. The highest and lowest score drop out and the remaining three scores are summed up and form the subjective performance score.

Judges can observe the objective performance measure before they must submit their subjective performance evaluation. This is in line with many organizational settings in which the superior knows the objective performance such as sales or profitability before his or her evaluation about subjective criteria such as leadership or communication takes place.

Most of the international judges have been (semi-) professional ski jumpers themselves, even though that is not mandatory. Judges must have at least three years of national experience and undergone trainings and certifications to be eligible to judge international competitions. The FIS does not pay Judges beyond reimbursements for their cost of travel and accommodation. Furthermore, the FIS instructs the national ski associations to select and nominate the judges for the respective events up to 2 years in advance. Thus, judges might have an incentive to favor fellow citizens in order to stand in a favorable light with their own national association (Zitzewitz, 2006).

However, a variety of features in this setting constrains favoritism of the evaluators' subjective assessment. First, the objective measure – observable before the evaluators must impute the subjective scores – mitigates the magnitude of favoritism. In addition, the two measures at least partly capture corresponding dimensions insofar as a very far jump is more likely with a good style and vice versa. Second, there are five evaluators assessing the same performance. Hence, exaggerations can be identified much easier since individual evaluations can be compared with other evaluations of the same performance. Third, the highest and lowest subjective evaluations are deleted, and only the aggregate of the remaining three evaluations go into the final score. Fourth, the FIS publishes individual assessments on its website after each competition, including the names and nationalities of the judges. Finally, a committee of the FIS reviews individual evaluations for each competition and sanctions favoritism (FIS, 2020, p. 5): "a working group of the FIS Sub-Committee will determine whether the Judge has acted objectively in the evaluation process without a national bias. They determine the so-called 'Judges Points' (preference for their own nation over other nations)." In addition, judges' evaluations should not deviate by more than 0.5 points in comparison to the mean of the remaining four evaluations in more than 20 percent of their evaluations. Otherwise, they might not be assigned for the World Cup in the following season.

4 Research design

Sample and variable Measurement

Our sample covers all events of the individual men's ski jumping World Cup of the five seasons from 2017 to 2022. We hand collected the data from the homepage of the International Ski Federation. In total, our sample contains 153 events from 27 different locations and hills in 12 different countries. Most events took place in Germany

(40), followed by Norway (18), Austria (15) and Poland (15). The events up to 22/02/2020 were held with an audience being present; most of the subsequent events were conducted without a physical audience present due to the Covid-19 pandemic. We include both jumps of athletes who participated in the first and second round of the event, so we usually observe 60 jumps per competition. However, five events were aborted after the first round due to bad weather conditions. We exclude jumps that resulted in a fall or where the athlete was disqualified. Our final sample covers 8,981 jumps of 151 athletes from 18 different countries. We focus in our analysis on the judges' subjective performance evaluation. As five judges evaluate each jump, we receive 44,905 judge-jump observations of 158 judges coming from 21 different countries. Table 1 provides characteristics about the frequencies of athletes, judges, and their nationalities.

[Please insert Table 1 here]

The variable *Style* is the subjective performance evaluation of judge *j* for jump *p* of athlete *a*. *Same nation* is an indicator variable that takes the value of one if the athlete has the same nationality as the judge. *Compatriot* is an indicator variable that takes the value of one if the athlete performing the jump is a compatriot of one of the four other judges on the panel. *Home* is an indicator variable that takes the value of one if the athlete performs the jump in his home country. *Audience* is an indicator variable that takes the value of one if the event takes place in front of an audience that is physically present.

We capture the objective performance of the jump with the variable *Distance*. It measures the distance points received for each jump. *Distance²* is the squared measure of distance points. We use the squared value of *Distance* to account for an increased difficulty of landing with a good style if the jump is longer. As the slope of the hill decreases in the lower landing area, it becomes more difficult to land with the appropriate style for very long jumps. *Distance meter* is the length of each jump in meters. In addition, our data set contains information about the point compensation for exogenous factors that influence the performance of the jump. The compensation depending on the starting gate of the jump is denoted as *Gate*, the compensation for the wind conditions is denoted as *Wind*.

[Please insert Table 2 here]

Table 2 shows descriptive statistics on the variables of interest. We report these data for the full sample as well as for different subsamples.

The subjective evaluation of style ranges from 6 to 20, with a mean of 17.82, and a standard deviation of 0.84. The in-group standard deviation is 0.29 indicating that the five evaluations of the same performance are close to each other. Indeed, for over 90 percent of the jumps, the minimum and maximum scores are just one point or less apart. In our sample, about 9 percent of evaluations are from a judge that has the same nationality as the athlete. In about 38 percent of evaluations, one of the other judges in the panel is a compatriot of the evaluated athlete. In addition, 13 percent of jumps are performed in the athletes' home country. Since athletes that jump in an event in their home country always have a compatriot on the panel, we distinguish between the full sample and a subsample without jumps in the athlete's home country (excl. Home) in our analysis. Finally, 71 percent of jumps are in front of a physically present audience.

The subjective evaluation *Style* is significantly higher if the judge shares the same nationality as the athlete. *Style* is also significantly higher if the athlete has a compatriot in the panel. The evaluation and jump take place in front of a physically present audience, and if the athlete performs in his home country (all four differences statistically significant at the 1 percent level). These descriptive statistics suggest that the judges' evaluations, the athletes' performances, or both tasks are associated with changes depending on these four indicator variables. However, due to the bivariate nature of this analysis, we postpone a more detailed discussion of the findings to the multivariate regression analyses.

Regression models

As a starting point, we estimate a regression model that investigates two different types of favoritism biases. First, a straightforward favoritism resulting from the judge being of the same nationality (a permanent in-group) as the athlete, i.e., direct favoritism. Second, a more subtle form of favoritism resulting from having a compatriot in the panel, i.e., indirect favoritism. This refers, for example, to an evaluation where a judge from Norway is judging a jumper from Austria and one of the other four judges in his judging panel is from Austria as well. Hence, we investigate if the characteristics of the judge's panel (a temporary in-group) influence his evaluation. To examine the existence of these two types of biases in our sample we run the following regression model based on Sandberg (2016)[9]:

$$Style_{j,a,p} = \beta_1 Same\ nation_{j,a,p} + \beta_2 Compatriot_{j,a,p} + \beta_3 Home_{j,a,p} + \eta_j + \gamma_{a,s} + \lambda_l + X_p + \varepsilon_p, \ (1)$$

where *Style* refers to the subjective performance evaluation from judge j, of jump p, performed by athlete a. η_j captures judge fixed effects, $\gamma_{a,s}$ captures athlete-per-season fixed effects, and λ_l captures location fixed effects. X_p is a vector that captures the objective performance indicators, namely the distance points *Distance*, its squared value *Distance²*, as well as the compensation for exogenous factors *Wind* and *Gate*.

The regression model is estimated by ordinary least squares (OLS) with robust standard errors clustered at the jump level. Based on prior literature, we expect the coefficient β_1, capturing direct favoritism by a fellow country man, to be positive. The empirical evidence for an indirect favoritism bias captured by β_2 is mixed. Zitzewitz (2006) finds a significant negative association between the subjective evaluation and having a compatriot in the panel using data from 23 events in ski jumping from the years 2001 and 2002. On the other hand, Krumer et al. (2021) do not find a significant result using a larger data sample of 203 events between the years 2010 and 2017. However, Sandberg (2018) shows a positive and significant indirect favoritism bias of judges in dressage competitions. Hence, we do not predict a certain sign for the coefficient β_2 in our regression models.

To examine Hypothesis H1 that the physical presence of a stakeholder is negatively associated with favoritism biases, we expand model (1). We include the indicator variable *Audience* as a main effect and as interaction effect with the three independent variables of interest, namely *Same nation*, *Compatriot*, and *Home*.

Hence, we run the following regression model:

$$Style = \beta_1 Same\ nation + \beta_2 Compatriot + \beta_3 Home + \beta_4 Audience + \beta_5 Same\ nation * Audience$$

[9] Since we are interested in the effect of a physically present audience, we cannot use a model with jump fixed effects, as this would absorb the variation in the physically present audience. Moreover, as outlined by Sandberg (2018), using a model with jump fixed effects in combination with an indicator variable if the evaluator and the athlete have the same nationality comes with a major drawback: In our setting, it would estimate the difference between the evaluation of a judge from the same nation as the athlete and the average evaluations of the other four judges. Hence, if the four other judges are influenced by the athlete's compatriot on the panel, the coefficient on the indicator would be distorted, as it measures the difference between both biases.

$$+\beta_6 Compatriot * Audience + \beta_7 Home * Audience + \eta_j + \gamma_{a,s} + \lambda_l + X_p + \varepsilon_p \quad (2)$$

Following Hypothesis H1, we expect a negative coefficient on the interaction terms of *Same nation* and *Audience,* β_5, as well as on *Compatriot* and *Audience*, β_6, suggesting that the physical presence of an audience is associated with lower favoritism in subjective performance evaluation.

5 Results

In the following, we present the regression results of our study. First, we present the baseline results with respect to the direct as well as the indirect favoritism bias. Second, we elaborate upon the main findings related to H1 (i.e., model (2)) and present robustness checks. Finally, we present additional analysis to investigate if the physical presence of an audience affects athletes in our setting as well.

Baseline results

We present the baseline results of model (1) in Table 3. First, we examine whether a direct favoritism bias is present. The estimated coefficient β_1 on *Same nation* as shown in the first column of Table 3 is positive and statistically significant (β_1 = 0.096, p-value < 0.01). Thus, the evaluation of an athlete's performance is on average 0.096 points higher if the athlete has the same nationality as the judge. The magnitude of this direct favoritism bias corresponds to 11.4 percent of the overall standard deviation and 33.7 percent of the in-group standard deviation of the subjective evaluations. Therefore, in line with our expectation and prior literature, we find evidence supporting a direct favoritism bias.

Second, we examine whether an indirect favoritism bias is present. The estimated coefficient β_2 on *Compatriot* is positive and statistically significant (β_2 =0.034, p-value < 0.05). Thus, the evaluation of an athlete's performance is on average 0.034 points higher if he has a compatriot on the panel. The magnitude of this indirect favoritism bias corresponds to 4 percent of the overall standard deviation and 11.9 percent of the in-group standard deviation. These findings indicate that an indirect favoritism bias exists in the sample.[10] Although the absolute magnitude of indirect favoritism is less than the

[10] The finding of an indirect favoritism bias in our sample seems to be explained by the observations without a physically present audience (see below). Presented in Appendix 2, we perform a sample split regarding the

magnitude of direct favoritism, it is not negligible because it can affect not just one but all four judges on the panel.

Third, we test if evaluators are biased in favor of athletes who perform in their home country, either because they want to favor local athletes, or because the crowd influences their evaluations through noise (Nevill et al., 2002; Unkelbach & Memmert, 2010). The coefficient β_3 on *Home* is positive but lacks statistical significance (β_3 =0.025, p-value = 0.21). Thus, we do not find statistically significant support for this prediction.

Taken together, an athlete who has a compatriot in the jury receives on average 0.232 more in total style points from all five judges, of which 0.096 points come from his compatriot and 0.136 points come from the remaining four judges. These results suggest that judges' subjective evaluations are biased when the athlete has the same nationality as the judge and when the athlete has a compatriot on the panel.

Athletes who jump in their home country always have a compatriot on the panel. So, our inferences would be distorted if athletes jump with a better style when performing in their home country beyond what is measured with the objective performance indicators.[11] Therefore, we run the regression for a subsample where we exclude jumps of athletes who jump in their home country (Column 2). For this subsample the coefficient on *Same nation* is lower but remains positive and statistically significant (β_1 = 0.089, p-value < 0.01). In addition, the coefficient on *Compatriot* is slightly higher and remains statistically significant (β_2 = 0.036, p-value < 0.05). These findings suggest that our inferences are not distorted by athletes that jump in their home country.

[Please insert Table 3 here]

For the control variables, we find that the distance of the jump matters. We find a positive and statistically significant (p-value < 0.01) association between *Distance* and *Style*. These findings suggest that the dimensions captured by the two measures for objective and subjective performance overlap. A jump performed in good style is more likely to go further down the hill, and a longer jump is more likely to be in good style.

indicator *Audience* and run model (1) on both subsamples. The coefficient on *Compatriot* is positive (0.077) and statistically significant (p-value < 0.05) for the observations without an audience on site. However, for the observations with a physically present audience it is positive (0.018) but not statistically significant, which corresponds to the results of Krumer et al. (2021).

[11] It could be the case that athletes train more often on hills in their home country and, therefore, are more accustomed to jumping on them.

We find a negative and statistically significant (p-value < 0.01) association between *Distance²* and *Style*. This result indicates that objective performance and subjective evaluation share an inversed u-shaped relation, highlighting that the further a jump goes, the harder it is to finish it with a good landing. Finally, we find a positive and statistically significant (p-value < 0.01) association between *Wind* and *Style* while we find a negative statistically significant (p-value < 0.05) association between *Gate* and *Style*.

Main results

In Table 4, we present our main findings. For the first two columns, we replicate the baseline-model including the indicator variable *Audience* solely as a main effect, once with hill-type fixed effects (1a) and once with location fixed effects (1b). The coefficients on *Audience* are positive, and statistically significant (p-value < 0.01). This suggests, that there is a positive association between the physically presence of an audience and subjective evaluations, i.e. judges on average award higher points in the presence of a live audience.

In the following columns (2a)-(3b) we present our main results related to Hypothesis H1, about the association of a physically present stakeholder and favoritism. For this, we focus on the interactions of *Same nation*Audience*, *Compatriot*Audience*, and *Home*Audience* as described in model (2). We run the fully interacted model with hill-type fixed effects in the columns a, and with location fixed effects in the columns b.

[Please insert Table 4 here]

First, we observe that the coefficient on *Same nation* increases to 0.166 (column (2a)), is positive and highly statistically significant (p-value < 0.01). Thus, the evaluation of an athlete's performance is on average 0.166 points higher if the judge evaluates a fellow citizen and when there is no audience on site.

Similarly, the coefficient on *Compatriot* increases to 0.109, is positive and statistically significant (p-value < 0.01). Thus, a judge scores an athlete's performance on average 0.109 points higher if the athlete has a compatriot on the judging panel and when there is no audience on site. These findings suggest that both types of favoritism are more pronounced when the evaluation takes place without an audience on site.

Second, our results suggest that there exists a negative association between judges' biases in subjective evaluations and the physical presence of a stakeholder. In particular, we find that the coefficient on the interaction of *Same nation*Audience* is negative (-0.085;

column (2a)) and statistically significant (p-value < 0.01). Thus, this finding indicates that the presence of an audience on site is associated with a lower direct favoritism bias by about 51 percent in our setting.

Similarly, we find that the coefficient on the interaction of *Compatriot*Audience* is negative (-0.089) and statistically significant (p-value < 0.01). This finding shows that the presence of an on-site audience is associated with a reduced indirect favoritism bias by about 80 percent in our setting.

When including location fixed effects instead of hill-type fixed effects the coefficients on *Same nation* and *Compatriot* are slightly smaller but remain statistically significant (column (2b)). In addition, the estimated coefficient on the interaction of *Same nation*Audience* in column (2b) is smaller (-0.075) but remains statistically significant (p-value < 0.05). Similarly, the coefficient on the interaction of *Compatriot*Audience* remains negative (-0.082) and remains statistically significance (p-value < 0.05).

As before, we estimate model (2) using the subsample of jumps not performed in the athlete's home country reported in column (3a) and (3b). The magnitude of the coefficient on the interaction of *Same nation*Audience* increases slightly (-0.092) and remains statistically significant (p-value < 0.01). Furthermore, the magnitude of the coefficient on the interaction of *Compatriot*Audience* increases (-0.094) and remains statistically significant (p-value < 0.01). When we run the full interacted model including location fixed effects (column 3b) the coefficients largely resemble those with hill type fixed effects.

Third, we concentrate on a possible home bias. We find a positive (0.038) and statistically significant coefficient (p-value < 0.10) on *Home* in our first model specification (column (1a)). However, the coefficient loses statistical significance when using location fixed effects (column (1b)) and when running the full interacted model (column (2a) and (2b)). Thus, our results only provide weak evidence in favor of a home bias, meaning that athletes receive higher scores when competing in their home country. Also, we observe that the coefficient on the interaction of *Home*Audience* is positive (0.052; column 2(a)), but not significantly different from zero (p-value = 0.26).

Overall, the results in Table 4 largely confirm Hypothesis 1, stating that a physically present stakeholder is associated with reduced favoritism in subjective evaluations.

Our results complement prior research on the influence of a physically present crowd on evaluators. Recent studies show that a physically present audience can pressure the evaluator's decisions in their favor (e.g., Garicano et al., 2005; Bryson et al., 2021). For instance, referees in football award significantly more extra time in matches where the home team is closely behind in order to satisfy the crowd (Garicano et al., 2005). Similarly, referees award fewer yellow cards to the away team without an audience on site (Bryson et al., 2021). While these studies show that a crowd can impose their preferences on evaluators, our results indicate that the presence of a crowd can mitigate the preferences of evaluators. A possible explanation is the different nature of the competitions. The two teams in football compete directly against each other, i.e., every decision by the referee in favor of one team is always a decision against the other team. In contrast, athletes in ski jumping compete more independently against each other. Thus, a higher evaluation for one athlete only indirectly harms the other athletes as it decreases their likelihood of having the highest score at the end.

It is important to remember that the events are broadcasted, and individual evaluations revealed on the internet ex-post for the entire period. Hence, the presented results allow for a clearer identification of the marginal effect of interest (i.e., from the physical presence of a stakeholder) holding other transparency mechanisms constant.

Robustness checks

In the next step, we perform robustness checks to validate our main findings. We run our main regressions for different subsamples. Table 5 presents the results of running model (2) on three different subsamples.[12] Again, we vary between hill-type fixed effects and location fixed effects.

[Please insert Table 5 here]

A possible distortion could result from only including the performances of the top 30 athletes who make it to the second round in our analyses. Since 50 athletes start in the first round, the worst 20 performances of the first round are excluded, while all performances of the second round are included. Thus, we run model (2) for a subsample in which we only include the top 25 performances of each round to exclude the worst

[12] In addition, we replicate our main regression analysis while changing the operationalization of the physically present audience. Results are presented in Appendix 3 and largely confirm our predictions.

performances of both rounds (column (1a) and (1b)). The results resemble our prior findings.

Next, we focus on the more recent seasons. Although ski jumping is an old sport with a long tradition, the sport continues to evolve due to innovations in material or changes in rules. For instance, compensation for wind and gate factors was introduced by the FIS only in 2009. Even though we are not aware of any serious rule changes during the sample period, we re-estimate model (2) for the last two seasons of our samples, which also makes the number of jumps with and without an audience more balanced. The results are presented in columns (2a) and (2b). The coefficients on *Same nation* and *Compatriot* slightly decrease in comparison to the ones from our main results but remain statistically significant. Similarly, the coefficients on the interactions of *Same nation*Audience* and of *Compatriot*Audience* remain negative and significant at least to the 5 percent level.

Finally, we seek to mitigate concerns that our results are driven by outliers or jumps where there is extreme disagreement about the true performance in style. Thus, we exclude jumps where the difference between the minimum and maximum subjective evaluation differs by more than two points. The results for this subsample (column (3a) & (3b)) are consistent those found previously.

Additional analysis

In the previous section, we focused on the effect of a physically present audience on evaluators while controlling for performance indicators. However, similar effects of the audience can also affect individual's performance (e.g., Triplett, 1898; Travis, 1925). Indeed, Scoppa (2021) provides evidence that playing without a supportive crowd during the Covid-19 pandemic reduced the home advantage across the top 5 leagues in European football. Thus, we examine associations between the athlete's objective performance and the presence of an audience on site. For this, we run the following regression model:

$$Distance\ meter_j = \beta_1 Home + \beta_2 Audience + \beta_3 Home * Audience + \gamma_{a,s} + \lambda_t + \varepsilon_p \qquad (3)$$

The dependent variable *Distance meter* is measured as the jump's distance in meters. Again, we control for performance indicators, namely wind points (*Wind*) as well as gate points (*Gate*) and include athlete-per-season fixed effect, $\gamma_{a,s}$, as well as location fixed effects, λ_l.

[Please insert Table 6 here]

For a better comparison with prior studies, we initially focus on the main effect of an athlete jumping in his home country for the full sample in column (1). In columns (2a) and (2b) we split the sample according to whether the jumps were performed with a physically present audience or not. For the full sample, the coefficient on *Home* is positive (0.53; column 1) but not statistically significant to conventional levels (p-value = 0.14). An athlete jumps on average 0.53 meters longer when competing in his home country. However, limiting the analysis to jumps performed with an audience on site, the coefficient on *Home* increases to 0.74 and becomes statistically significant (p-value < 0.10), which corresponds to the results of Krumer et al. (2021). The coefficient on *Home* is negative (-0.39) and lacks statistically significant (p-value = 0.36) if we only include jumps without an audience on site. These findings suggest that athletes have a home advantage under 'normal' circumstances.

Finally, we present the results of the full model (3) in column (3). The coefficient on *Home* decreases to 0.07 and becomes statistically insignificant from zero. The coefficient on *Audience* is positive (1.80) and statistically significant (p-value < 0.1). Thus, a jump performed in front of a physically present audience is on average 1.80 meters longer. If a home advantage in ski jumping is driven by the cheering of the home crowd[13], we would expect a positive coefficient on the interaction of *Home*Audience.* While the direction of this coefficient 0.67 is in line with this prediction, it lacks statistical significance (p-value = 0.25).

In sum, we provide slight evidence for two conclusions. First, a physically present audience on average increases athletes' performance in the objective performance dimension. Second, athletes in our setting have a slight home advantage under 'normal' circumstances.

---

[13] There exists an extensive literature on the home advantage in sports (e.g., Nevill & Holder, 1999). Three explanations are proposed in the literature: audience support, familiarity with the facilities, and travel fatigue.

6 Discussion

Implications of our findings

Our study focuses on individual's subjective evaluations in committees, which is a ubiquitous element across organizations. For instance, the highest decision-making body in corporations, the board of directors, regularly takes (subjective) evaluations of projects, investments, or products that shape corporate performance. Clearly, our setting has some unique features, such as no communication among evaluators, simultaneous voting, and a short evaluation process that differ from traditional settings. However, due to the advantages of this setting, in particular the clear incentive structure, the high stakes, and the competitive environment, we see it as a good setting to draw conclusions for organizations.

First, our results have important implications for organizations in their quest to reduce favoritism in evaluation decisions by individuals as well as by groups of individuals. Organizations often delegate decision making to more than one individual to mitigate both potential favoritism biases that one member of the evaluation panel might have, as well as to reduce consequences of such biases for the final decision. For instance, in academic and corporate settings more than one individual is mostly in charge to assess candidates in appointment or promotion committees. Our findings show that the presence of an athlete's compatriot in the panel is positively associated with the subjective evaluation by the other panel members. Similarly, appointment or promotion committees usually include one of the candidate's superiors or confidants, who likely have a biased assessment in favor of the candidate.

Since this indirect favoritism bias might potentially influence the other committee members, our results implicate that the expansion of the panel to reduce the favoritism of one member might have a contrary effect. Considering the interactive nature of those decisions, and the fact that panel members are likely to continue to work together in the future might reinforce the effect.

Interestingly, our findings also offer a possible remedy to this problem: Increase transparency during the decision-making process. Organizations can achieve this by including and appointing representatives of various stakeholders to the committee. Even if those representatives have no vote on the final decision, their physical presence during the decision-making process helps to arrive at a fairer and less biased evaluation.

Similarly, organizations regularly organize innovation or cost-saving competitions, as mentioned earlier. Based on our findings, the final pitches of employees' ideas should rather take place in town-hall meetings in front of the workforce than only before the jury members to ensure fairer evaluations. In sum, we argue that organizational designers should form practices and policies of committees in a way that does not only increase transparency and accountability ex-post, but also during the evaluation process.

Second, our findings also have broader implications. During the Covid-19 pandemic, many organizations conduct their general meeting virtual instead of in person. While this certainly helps in breaking the chains of infection, it might do so with an underappreciated cost. The physical presence not only of the corporation's shareholders, but also of other stakeholders like employees, auditors or the press leads to an increased level of transparency.[14] Following our study's train of thought, this increased transparency might mitigate managers' inherent preferences. Hence, managers might provide a less biased evaluation about the organization's position and outlooks when the general meeting takes place in person instead of virtually. Indeed, Brochet et al. (2021) find that virtual meetings are shorter, less likely to include a business presentation, and when they do the language is more generic.

Finally, the 'debiasing' of information is not only a central topic in managerial accounting, but also in financial accounting. For instance, analysts or rating agencies face a comparable incentive structure as the judges in our setting. On the one hand, sell-side analysts have incentives to bias their forecasts to improve the relationship with organization's management (Francis & Philbrick, 1993). On the other hand, analyst have strong career concerns and providing accurate forecasts serves them in the short and long term (Hong & Kubik, 2003). So, increasing transparency of how analysts derive at their recommendations can help in reducing the 'earnings guidance game' of overly optimistic forecasts.

7 Conclusion

We investigate if a subtle feature of transparency – the physical presence of a stakeholder during the evaluation process – reduces biases in subjective evaluations

[14] Activist J. Chevedden raised a similar concern in a letter against General Motor's decision to hold the annual conference virtually (PX14A6G Filing, May 2, 2019): "An in-person annual meeting is a motivator of good performance by management and directors. Who wants to stand in front of a live audience and explain shrinking sales, epic recalls and loss of market share? It is so much easier to explain it to a microphone."

caused by preferences to favor in-group members. To address this question empirically, we use data from professional ski jumping, where professional judges subjectively evaluate athletes' performances. In particular, we examine if a physically present audience moderates favoritism biases in the subjective evaluation by exploiting the natural experiment that sports events where hold without an on-site audience during the Covid-19 pandemic.

We derive three main findings. First, we find that subjective evaluations in our setting suffer from two biases: (1) Direct favoritism bias, occurring when the judge evaluates an athlete, who has the same nationality; (2) Indirect favoritism bias, occurring when the judge evaluates an athlete whose compatriot is also on the judging panel. Second, we provide evidence that the physical presence of a stakeholder mitigates both types of favoritism. More specifically, we find that the direct favoritism bias is about 50 percent lower, and the indirect favoritism bias becomes statistically insignificant if there is an audience physically present at the event. Based on social facilitation theory, we argue that the physical presence of a stakeholder raises evaluators' arousal, which increases the salience of unbiased evaluations. Thus, evaluators face higher cost of considering their preference to favor members of the evaluator's in-group.

Lastly we find that subjective performance evaluations lead evaluators on average to award higher evaluation levels. This suggests that direct observation of the evaluation lead evaluators to make sure their evaluation is not considered to be too low.

We contribute to the accounting literature on subjective performance evaluation (Gibbs et al, 2004; Bol, 2008, 2011). We extend prior studies by showing that favoritism not only emerges directly through evaluators' own preferences, but also indirectly through the composition of the decision-making body. In addition, we provide evidence on how biases in the subjective evaluation can be mitigated through the physical presence of a stakeholder. Thus, we highlight a technique that helps in debiasing subjective evaluations.

Also, we contribute to the literature on the effects of social factors (Akerlof, 1980; Bernheim, 1994; Becker & Murphy, 2000). Prior research shows that a crowd can influence a decision-maker in accordance with their preferences (e.g., Garicano et al., 2005; Bryson et al., 2021). We complement this research by highlighting a benefit of having a crowd physically present, as this reduces consequences of decision makers' preferences.

Our study is subject to some limitations. A potential correlation between the nationality of the judge and the relative quality of the athletes from that same country, which goes beyond what the objective performance indicators capture, would limit our inferences. In addition, our results are subject to the caveat of being based on a single industry, which might hinder the generalization to other settings. Professional sports have many unique characteristics, such as athletes' short careers, their youth, or sports-specific regulations (e.g., Black & Vance, 2021). However, several characteristics of our setting, such as the readily available objective as well as subjective performance measure, a strong competitive environment, the high stakes, and a variety of mechanisms to facilitate fair and unbiased evaluations suggest that inherent preferences should be muted in comparison to more traditional empirical settings. This implies that our findings might serve as a lower bound for estimates in typical organizational settings.

Overall, our main emphasis is on individuals' subjective evaluations and not on the actions of the athletes themselves. These individuals face a general task of subjectively evaluating the agents' performance with or without the physical presence of a stakeholder, as is the case in diverse organizational settings.

Appendix

Appendix 1

Variable Definitions

Variable name	Variable definitions
Audience	is an indicator variable that takes the value of 1 if the event takes place in front of a physically present audience.
Audience Size	is a proxy variable for the size of the audience. Its value is the capacity of the stadium for events that took place with a physically present audience and its value is 0 if the event took place without a physically present audience.
Compatriot	is an indicator variable that takes the value of 1 if the athlete has a judge in the panel with the same nationality.
Distance	measures the distance points received for each jump.
Distance²	is the squared measure of distance points.
Distance meter	is the length of each jump in meters.
Gate	captures the point compensation depending on the starting gate for each jump.
Home	is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country.
Same nation	is an indicator variable that takes the value of 1 if the athlete has the same nationality as the judge.
Style	denotes the subjective performance evaluation of a judge.
Wind	captures the point compensation dependent on the wind conditions for each jump.

Notes: Appendix 1 lists the variables used in the empirical analysis and their description.

Appendix 2

	Style	*Style*
	(1)	(2)
Same nation	0.083***	0.136***
	(0.018)	(0.035)
Compatriot	0.018	0.077**
	(0.017)	(0.035)
Home	0.033	-0.016
	(0.023)	(0.041)
Controls	Yes	Yes
Athlete-per-season FE	Yes	Yes
Judge FE	Yes	Yes
Location FE	Yes	Yes
Sample	with Audience	no Audience
N	32,030	12,875
Adj. R^2	0.47	0.48

Notes: All estimates are based on OLS regressions with fixed effects as denoted in the table. The dependent variable is ***Style*** and measures the judges' subjective performance evaluation. ***Same nation*** is an indicator variable that takes the value of 1 if the athlete has the same nationality as the judge. ***Compatriot*** is an indicator variable that takes the value of 1 if the athlete has a judge in the panel with the same nationality. ***Home*** is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country. ***Audience*** is an indicator variable that takes the value of 1 if the event takes place in front of a physically present audience. We control for additional performance indicators, namely ***Distance***, ***Distance²***, ***Wind*** and ***Gate.*** Robust standard errors are clustered at the jump level. ***, **, * denote significance at the 1%, 5%, 10%-level, respectively.

Appendix 3

Robustness test with different operationalization of a physical present audience						
	Style	*Style*	*Style*	*Style*	*Style*	*Style*
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Same nation	0.100***	0.097***	0.162***	0.150***	0.159***	0.148***
	(0.016)	(0.015)	(0.028)	(0.029)	(0.031)	(0.031)
Audience Size	0.011***	0.004*	0.014***	0.007**	0.014***	0.008***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
*Same nation*Audience Size*			-0.008**	-0.007**	-0.008**	-0.008**
			(0.003)	(0.003)	(0.003)	(0.003)
Compatriot	0.0441***	0.034**	0.103***	0.092**	0.113***	0.096**
	(0.015)	(0.015)	(0.028)	(0.029)	(0.030)	(0.030)
*Compatriot*Audience Size*			-0.008**	-0.008***	-0.009*	-0.008
			(0.003)	(0.003)	(0.003)	(0.003)
Home	0.038*	0.025	-0.003	-0.026		
	(0.020)	(0.020)	(0.040)	(0.041)		
*Home*Audience Size*			0.005	0.007		
			(0.004)	(0.004)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Athlete-per-season FE	Yes	Yes	Yes	Yes	Yes	Yes
Judge FE	Yes	Yes	Yes	Yes	Yes	Yes
Hill type FE	Yes	No	Yes	No	Yes	No
Location FE	No	Yes	No	Yes	No	Yes
Sample	Full	Full	Full	Full	excl. Home	excl. Home
N	44,905	44,905	44,905	44,905	39,135	39,135
Adj. R^2	0.45	0.46	0.45	0.46	0.45	0.46

Notes: All estimates are based on OLS regressions with fixed effects as denoted in the table. The dependent variable is ***Style*** and measures the judges' subjective performance evaluation. The variable ***Audience Size*** is measured as the natural logarithm of the capacity at the location for the event with a physically present audience and zero otherwise. ***Same nation*** is an indicator variable that takes the value of 1 if the athlete has the same nationality as the judge. ***Compatriot*** is an indicator variable that takes the value of 1 if the athlete has a judge in the panel with the same nationality. ***Home*** is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country. We control for additional performance indicators, namely ***Distance***, ***Distance²***, ***Wind*** and ***Gate***. Robust standard errors are clustered at the jump level. ***, **, * denote significance at the 1%, 5%, 10%-level, respectively.

References

Aiello, J. R., & Douthitt, E. A. (2001). Social facilitation from Triplett to electronic performance monitoring. *Group Dynamics: Theory, Research, and Practice*, 5(3), 163–180.

Akerlof, G. A. (1980). A Theory of Social Custom, of Which Unemployment May be One Consequence. *The Quarterly Journal of Economics*, 94(4), 749.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics*, 127(2), 1017–1055.

Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective Performance Measures in optimal Incentive Contracts. *The Quarterly Journal of Economics*, 109(4), 1125–1156.

Becker, G., & Murphy, K. (2009). Social economics. Harvard University Press.

Bernheim, B. D. (1994). A Theory of Conformity. *Journal of Political Economy*, 102(5), 841–877.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3(1), 27–52.

Black, D. E., & Vance, M. D. (2021). Do first impressions last? The impact of initial assessments and subsequent performance on promotion decisions. *Management Science*, 67(7), 4556–4576.

Bol, J. C. (2008). Subjectivity in Compensation Contracting. *Journal of Accounting Literature*, 23529(27), 1–27.

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5), 1549–1575.

Bol, J. C., & Smith, S. D. (2011). Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review*, 86(4), 1213–1230.

Brochet, F., Chychyla, R., & Ferri, F. (2021). Virtual Shareholder Meetings. Working Paper.

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D., & Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during Covid-19. *Economics Letters*, 198, 109664.

Chapman, B. K. S., & Southwick, L. (1991). Testing the Matching Hypothesis: The Case of Major-League Baseball. *The American Economic Review*, 81(5), 1352–1360.

Cottrell, N. B., Wack, D., Rittle, R. H., & Sekerak, G. J. (1968). Social Facilitaion of Dominant Responses by the Presence of an Audience and the mere Presence of Others. *Journal of Personality and Social Psychology*, 9(3), 245–249.

Demeré, B. W., Sedatole, K. L., & Woods, A. (2019). The role of calibration committees in subjective performance evaluation systems. *Management Science*, 65(4), 1562–1585.

Du, F., Tang, G., & Mark Young, S. (2012). Influence activities and favoritism in subjective Performance Evaluation: Evidence from Chinese state-owned enterprises. *The Accounting Review*, 87(5), 1555–1588.

Eyring, H., Ferguson, P. J., & Koppers, S. (2021). Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment. *Journal of Accounting Research*, 59(2), 657–711.

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, 833-860.

Fehrenbacher, D. D., Schulz, A. K. D., & Rotaru, K. (2018). The moderating role of decision mode in subjective performance evaluation. *Management Accounting Research*, 41, 1–10.

Ferguson, P. J. (2021). Do managers distinguish between effects of luck and effort on employees' performance outcomes? Evidence from a high-stakes field setting. Working Paper.

Financial Times. (2020). Does home advantage exist without football's partisan fans?

FIS. (2018). The International Ski Competition Rules (ICR). Book III Ski Jumping.

FIS. (2020). Guidelines for FIS Jumping Judges.

Fisman, R., Paravisini, D., & Vig, V. (2017). Cultural proximity and loan outcomes. *The American Economic Review*, 107(2), 457–492.

Francis, J., & Philbrick, D. (1993). Analysts' Decisions As Products of a Multi-Task Environment. *Journal of Accounting Research*, 31(2), 216.

Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under social pressure. *Review of Economics and Statistics*, 87(2), 208–216.

Gibbs, M., Merchant, K. A., Van Der Stede, W. A., & Vargus, M. E. (2004). Determinants and effects of subjectivity in incentives. *The Accounting Review*, 79(2), 409–436.

Grabner, I., Künneke, J., & Moers, F. (2020). How calibration committees can mitigate performance evaluation bias: An analysis of implicit incentives. *The Accounting Review*, 95(6), 213–233.

Gradwohl, R. & Feddersen, T. (2018). Persuasion and Transparency. *Journal of Politics,* 80(3), 903-915.

Hong, H., & Kubik, J. D. (2003). Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts. *Journal of Finance*, 58(1), 313–351.

Krumer, A., Otto, F., & Pawlowski, T. (2021). Nationalistic Bias among International Experts: Evidence from Professional Ski Jumping. *The Scandinavian Journal of Economics*.

Levit, D., & Malenko, N. (2016). The Labor Market for Directors and Externalities in Corporate Governance. *Journal of Finance*, 71(2), 775–808.

Levy, G. (2007). Decision making in committees: Transparency, reputation, and voting rules. *The American Economic Review*, 97(1), 150–168.

Moers, F. (2005). Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society*, 30(1), 67–80.

Mullen, B., Bryant, B., & Driskell, J. E. (1997). Presence of others and arousal: An integration. *Group Dynamics*, 1(1), 52–64.

Murphy, K. J., & Oyer, P. (2005). Discretion in Executive Incentive Contracts: Theory and Evidence.

Nevill, A. M., Balmer, N. J., & Mark Williams, A. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4), 261–272.

Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport. An overview of studies on the advantage of playing at home. *Sports Medicine*, 28(4), 221–236.

Ogura, Y., & Matsushima, T. (2011). Social facilitation revisited: Increase in foraging efforts and synchronization of running in domestic chicks. *Frontiers in Neuroscience*, 5, 91.

Parsons, C. A., Sulaeman, J., Yates, M. C., & Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *The American Economic Review*, 101(4), 1410–1435.

Pessin, J. (1933). The Comparative Effects of Social and Mechanical Stimulation on Memorizing. *The American Journal of Psychology*, 45(2), 263–270.

Pessin, J., & Husband, R. W. (1933). Effects of social stimulation on human maze learning. *Journal of Abnormal and Social Psychology*, 28(2), 148–154.

Prendergast, C. (1999). The Provision of Incentives in Firms. In *Journal of Economic Literature,* 37(1), 1–63.

Prendergast, C., & Topel, R. H. (1996). Favoritism in Organizations. *Labour Economics*, 104(5), 958–978.

Price, J., & Wolfers, J. (2010). Racial discrimination among NBA referees. The *Quarterly Journal of Economics*, 125(4), 1859–1887.

Sandberg, A. (2018). Competing Identities: A Field Study of In-group Bias Among Professional Evaluators. *Economic Journal*, 128(613), 2131–2159.

Scholten, H., Schneemann, S., & Deutscher, C. (2020). The Impact of Age on Nationality Bias and Cultural Proximity Bias: Evidence from Ski-Jumping. *Journal of Institutional and Theoretical Economics*, 176(4), 1-27.

Scoppa, V. (2021). Social Pressure in the Stadiums: Do Agents Change Behavior without Crowd Support? *Journal of Economic Psychology*, 82, 102344.

Shayo, M., & Zussman, A. (2011). Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics*, 126(3), 1447–1484.

The Economist. (2020). Empty stadiums have shrunk football teams' home advantage.

Travis, L. E. (1925). The effect of a small audience upon eye-hand coordination. *Journal of Abnormal Psychology and Social Psychology*, 20(2), 142–146.

Triplett, N. (1898). The Dynamogenic Factors in Pacemaking and Competition. *The American Journal of Psychology*, 9(4), 507–533.

Unkelbach, C., & Memmert, D. (2010). Crowd noise as a cue in referee decisions contributes to the home advantage. *Journal of Sport and Exercise Psychology*, 32(4), 483–498.

Yermack, D. (2004). Remuneration, Retention, and Reputation Incentives for Outside Directors. *The Journal of Finance*, 59(5), 2281–2308.

Zajonc, R. B., Heingartner, A., & Herman, E. M. (1969). Social enhancement and impairment of performance in the cockroach. *Journal of Personality and Social Psychology*, 13(2), 83–92.

Zitzewitz, E. (2006). Nationalism in winter sports judging and its lessons for organizational decision making. *Journal of Economics and Management Strategy*, 15(1), 67–99.

Table 1: Frequencies of Athletes, Judges, Jumps and Events

Percent	Freq.	No. of athletes	Nationality	No. of judges	Freq.	Percent
16.38	1,471	15	**GER**	25	7,164	15.95
13.94	1,252	15	**POL**	9	3,765	8.38
15.58	1,399	25	**AUT**	16	5,077	11.31
14.44	1,297	18	**NOR**	13	4,829	10.75
13.72	1232	18	**SLO**	12	3,045	6.78
10.90	979	12	**JPN**	10	1,439	3.2
4.39	394	6	**SUI**	10	3,025	6.74
3.17	285	9	**RUS**	6	2,368	5.27
2.47	222	7	**CZE**	8	1,677	3.73
1.74	156	8	**FIN**	10	2,960	6.59
0.78	70	2	**CAN**	3	478	1.06
0.61	55	1	**BUL**	0	0	0
0.64	47	2	**FRA**	5	1,442	3.21
0.55	49	5	**USA**	5	966	2.15
0.37	33	4	**ITA**	4	2,302	5.13
0.33	30	2	**EST**	2	244	0.54
0.09	8	1	**KAZ**	3	836	1.86
0.02	2	1	**TUR**	0	0	0
0.00	0	0	**KOR**	2	508	1.13
0.00	0	0	**ROU**	8	1255	2.79
0.00	0	0	**SVK**	3	624	1.39
0.00	0	0	**SWE**	4	659	1.47
0.00	0	0	**CHN**	1	242	0.54
100.0	8,981	151	**Total**	158	44,905	100

Table 2: Descriptive Statistics

Panel A: Summary statistics on the jump level

Variable		Obs.	Mean	SD	Min	Max
Home		8,981	0.13	0.33	0	1
Audience		8,981	0.71	0.45	0	1
Distance		8,981	77.88	27.38	-33.6	182.4
Distance²		8,981	6815.29	5544.83	17.6	33269.8
Distance meter		8,981	137.67	32.52	71.0	252.0
	Audience	6,406	140.53	35.32	71	252
	No Audience	2,575	132.09	23.79	80	249.5
Wind		8,981	-0.42	8.87	-30.2	37.8
Gate		8,981	0.68	4.19	-24.3	27.5

Panel B: Summary statistics on the evaluation level

Variable		Obs.	Mean	SD	Min	Max
Style		44,905	17.82	0.84	6	20
	Same nation	4,226	17.92	0.80	10	20
	Other nation	40,639	17.81	0.85	6	20
	Compatriot	16,874	17.87	0.81	9	20
	No Compatriot	28,031	17.79	0.87	6	20
	Home	5,770	17.89	0.83	9.5	20
	Away	39,135	17.81	0.85	6	20
	Audience	32,030	17.83	0.85	6	20
	No Audience	12,875	17.81	0.84	7	20
Same nation		44,905	0.10	0.29	0	1
Compatriot		44,905	0.38	0.48	0	1

Notes: This table lists the variables used in the empirical analysis and the corresponding distribution parameters. **Style** denotes the subjective performance evaluation of judges; ***Same nation*** is an indicator variable that takes the value of 1 if the athlete has the same nationality as the judge. ***Compatriot*** is an indicator variable that takes the value of 1 if the athlete has a judge in the panel with the same nationality. ***Home*** is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country. ***Audience*** is an indicator variable that takes the value of 1 if the event takes place in front of a physically present audience. ***Distance*** measures the distance points received for each jump. ***Distance²*** is the squared measure of distance points. ***Distance meter*** is the length of each jump in meters. ***Wind*** captures the point compensation dependent on the wind conditions for each jump. ***Gate*** captures the point compensation depending on the starting gate for each jump.

Table 3: Baseline Results

	Style	*Style*
	(1)	(2)
Same nation	0.096***	0.089***
	(0.016)	(0.017)
Compatriot	0.034**	0.036**
	(0.015)	(0.016)
Home	0.025	
	(0.020)	
Distance	0.060***	0.057***
	(0.002)	(0.002)
Distance²	-0.0001***	-0.0001***
	(0.00001)	(0.00001)
Wind	0.006***	0.006***
	(0.001)	(0.001)
Gate	-0.003*	-0.003**
	(0.002)	(0.002)
Athlete-per-season FE	Yes	Yes
Judge FE	Yes	Yes
Location FE	Yes	Yes
Sample	Full	Excl. Home
N	44,905	39,135
Adj. R^2	0.46	0.45

Notes: All estimates are based on OLS regressions with fixed effects as denoted in the table. The dependent variable is ***Style*** and measures the judges' subjective performance evaluations. ***Same nation*** is an indicator variable that takes the value of 1 if the athlete has the same nationality as the judge. ***Compatriot*** is an indicator variable that takes the value of 1 if the athlete has a judge in the panel with the same nationality. ***Home*** is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country. ***Distance*** measures the distance points received for each jump. ***Distance²*** is the squared measure of distance points. ***Wind*** captures the point compensation dependent on the wind conditions for each jump. ***Gate*** captures the point compensation depending on the starting gate for each jump. Robust standard errors are clustered at the jump level. ***, **, * denote significance at the 1%, 5%, 10%-level, respectively.

Table 4: Main Results

	Style	*Style*	*Style*	*Style*	*Style*	*Style*
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Same nation	0.100***	0.097***	0.166***	0.155***	0.165***	0.153***
	(0.016)	(0.016)	(0.029)	(0.029)	(0.031)	(0.032)
Audience	0.074***	0.044*	0.109***	0.075**	0.118***	0.081***
	(0.023)	(0.027)	(0.027)	(0.054)	(0.027)	(0.031)
*Same nation*Audience*			-0.085***	-0.075**	-0.092***	-0.083**
			(0.032)	(0.033)	(0.035)	(0.035)
Compatriot	0.041***	0.034*	0.109***	0.097***	0.117***	0.099***
	(0.015)	(0.015)	(0.029)	(0.030)	(0.030)	(0.043)
*Compatriot*Audience*			-0.089***	-0.082**	-0.094***	-0.080**
			(0.033)	(0.033)	(0.034)	(0.046)
Home	0.038*	0.025	-0.001	-0.025		
	(0.020)	(0.020)	(0.039)	(0.039)		
*Home*Audience*			0.052	0.068		
			(0.046)	(0.045)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Athlete-per-season FE	Yes	Yes	Yes	Yes	Yes	Yes
Judge FE	Yes	Yes	Yes	Yes	Yes	Yes
Hill-type FE	Yes	No	Yes	No	Yes	No
Location FE	No	Yes	No	Yes	No	Yes
Sample	Full	Full	Full	Full	excl. Home	excl. Home
N	44,905	44,905	44,905	44,905	39,135	39,135
Adj. R^2	0.45	0.46	0.45	0.46	0.45	0.46

Notes: All estimates are based on OLS regressions with fixed effects as denoted in the table. The dependent variable is ***Style*** and measures the judges' subjective performance evaluation. ***Same nation*** is an indicator variable that takes the value of 1 if the athlete has the same nationality as the judge. ***Compatriot*** is an indicator variable that takes the value of 1 if the athlete has a judge in the panel with the same nationality. ***Home*** is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country. ***Audience*** is an indicator variable that takes the value of 1 if the event takes place in front of a physically present audience. We control for additional performance indicators, namely ***Distance***, ***Distance²***, ***Wind*** and ***Gate***. Robust standard errors are clustered at the jump level. ***, **, * denote significance at the 1%, 5%, 10%-level, respectively.

Table 5: Robustness Test

	Style	*Style*	*Style*	*Style*	*Style*	*Style*
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Same nation	0.149***	0.138***	0.148***	0.141***	0.160***	0.148***
	(0.030)	(0.030)	(0.029)	(0.030)	(0.028)	(0.028)
Audience	0.112***	0.081***	0.148**	0.065*	0.115***	0.079***
	(0.028)	(0.030)	(0.030)	(0.035)	(0.025)	(0.027)
*Same nation*Audience*	-0.079**	-0.70**	-0.093**	-0.079**	-0.084***	-0.072**
	(0.034)	(0.034)	(0.038)	(0.038)	(0.032)	(0.032)
Compatriot	0.095***	0.083***	0.099***	0.090***	0.101***	0.089***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.028)	(0.029)
*Compatriot*Audience*	-0.084**	-0.078**	-0.108***	-0.096**	-0.085***	-0.078**
	(0.034)	(0.034)	(0.037)	(0.038)	(0.032)	(0.032)
Home	0.035	0.004	0.004	-0.000	0.001	0.025
	(0.032)	(0.032)	(0.039)	(0.039)	(0.039)	(0.039)
*Home*Audience*	0.029	0.054	0.021	0.017	0.060	0.077
	(0.039)	(0.039)	(0.061)	(0.060)	(0.045)	(0.044)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Athlete-per-season FE	Yes	Yes	Yes	Yes	Yes	Yes
Judge FE	Yes	Yes	Yes	Yes	Yes	Yes
Hill-type FE	Yes	No	Yes	No	Yes	Yes
Location FE	No	Yes	No	Yes	No	No
Sample	TOP25	TOP25	Last2Seasons	Last2Seasons	Diff<2.5	Diff<2.5
N	37,485	37,485	22,905	22,905	44,775	44,775
Adj. R^2	0.37	0.39	0.45	0.46	0.48	0.49

Notes: All estimates are based on OLS regressions with fixed effects as denoted in the table. The dependent variable is ***Style*** and measures the judges' subjective performance evaluation. ***Same nation*** is an indicator variable that takes the value of 1 if the athlete has the same nationality as the judge. ***Compatriot*** is an indicator variable that takes the value of 1 if the athlete has a judge in the panel with the same nationality. ***Home*** is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country. ***Audience*** is an indicator variable that takes the value of 1 if the event takes place in front of a physically present audience. We control for additional performance indicators, namely ***Distance***, ***Distance²***, ***Wind*** and ***Gate***. Robust standard errors are clustered at the jump level. ***, **, * denote significance at the 1%, 5%, 10%-level, respectively.

Table 6: Additional Analyses

	Distance meter	*Distance meter*	*Distance meter*	*Distance meter*
	(1)	(2a)	(2b)	(3)
Home	0.53	0.74*	-0.39	0.07
	(0.34)	(0.38)	(0.42)	(0.48)
Audience				1.80*
				(0.94)
*Home*Audience*				0.67
				(0.57)
Wind	-0.35***	-0.40***	-0.35***	-0.36***
	(0.059)	(0.07)	(0.054)	(0.06)
Gate	-0.01	-0.01	-0.01	-0.01
	(-0.05)	(0.05)	(0.08)	(0.04)
Athlete-per-season FE	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes
Sample	Full	with Audience	no Audience	Full
N	8,981	6,406	2,575	8,981
Adj. R^2	0.96	0.96	0.94	0.96

Note: The dependent variable ***Distance meter*** is the length of a given jump in meter. ***Home*** is an indicator variable that takes the value of 1 if the athlete performs the jump in his home country. ***Audience*** is an indicator variable that takes the value of 1 if the event takes place in front of a physically present audience. Robust standard errors are clustered at the location-level and presented in parentheses. ***, **, * denote significance at the 1%, 5%, 10%-level, respectively.