How trust in fairness and national identity affect the behavior of spectators and judges in sports

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

M.Sc. Felix Otto

Tübingen

2022

1. Betreuer:	Professor Dr. Tim Pawlowski
2. Betreuer:	Professor Dr. Ansgar Thiel

Tag der mündlichen Prüfung:	15.08.2022
Dekan:	Prof. Dr. Josef Schmid
1. Gutachter:	Professor Dr. Tim Pawlowski
2. Gutachter:	Professor Dr. Christian Deutscher

How trust in fairness and national identity affect the behavior of spectators and judges in sports

Dissertation

for the acquisition of the doctoral degree

from the Faculty of Economics and Social Sciences

at the University of Tübingen

Submitted by

M.Sc. Felix Otto

Tübingen

2022

Date of oral defense: 15.08.2022

Dean: Professor Dr. Josef Schmid

First reviewer: Professor Dr. Tim Pawlowski

Second reviewer: Professor Dr. Christian Deutscher

Acknowledgements

First, I would like to thank my supervisor Prof. Dr. Tim Pawlowski for providing me the opportunity to explore the field of (behavioral) sports economics. His guidance and support throughout my time as a PhD student allowed me to develop my interests and competencies as an academic researcher. I would also like to thank him for teaching and encouraging me to always keep trying to improve the quality of my work and to become a better researcher overall. Next, I would like to express my gratitude to my other co-authors. Georgios Nalbantis supported me in every aspect of my academic career and helped me with valuable discussions, great ideas, daily walks, and his excellent knowledge in the field of sports economics. Alex Krumer taught me the value and meaning of economic thinking in scientific research and beyond. Sonja Utz helped me to gain further insights into psychological research and related fields, which allowed me to expand my research horizon. Furthermore, I would like to thank my current and former colleagues of the Department of Sport Economics, Sport Management and Media Research, Verena Burk, Marcel Fahrner, Ute Schüttoff, Tim Wallrafen, Travis Richardson, Christina Tölle, and Paula Fischer, for their daily cooperation and support. Finally, I want to express my deep gratitude to my family and close relatives. They supported me throughout my personal and academic journey with much love and encouragement.

Table of contents

Abbreviations

ACS	American Community Survey
AIC	Akaike Information Criterion
ATE	Average treatment effect
ATENT	Average treatment effect on the non-treated
ATET	Average treatment effect on the treated
BIC	Bayesian Information Criterion
BSE	Bovine spongiform encephalopathy
CHF	Swiss francs
CIA	Conditional independence assumption
CMP	Conditional (recursive) mixed process
COO	Country of origin
CPI	Corruption perceptions index from Transparency International
CVM	Contingent valuation method
EAC	European Athletics Championships
FIS	Fédération Internationale de Ski
GAA	German Athletics Association
HS	Hill sizes
IAAF	International Association of Athletics Federations
IRR	Incidence rate ratios
ISTAF	Internationale Stadionfest
ITC	Intention-to-consume
IV	Instrumental variable
MLB	Major League Baseball
MLS	Major League Soccer
NAASE	North American Association of Sports Economists
NBM	Negative binomial model
NFL	National Football League
OLS	Ordinary least squares
OR	Odds ratios
PED	Performance-enhancing drugs
T&F	Track and field
TV	Television
UK	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States (of America)
WTP	Willingness-to-pay
ZIP	Zero-inflated Poisson

How trust in fairness and national identity affect the behavior of spectators and judges in sports

1 Introduction

This dissertation focuses on the role of social concerns i.e., *fairness* and *identity concerns*, for the economic behavior of two key agents in professional sports, i.e., *sport consumers* – hereinafter referred to as spectators – and *sport judges*. In this regard, the problem statement and focus of this dissertation are explained in Subchapter 1.1, followed by the presentation of the dissertation's structure in Subchapter 1.2.

1.1 Problem statement

Advances in behavioral economic research led to the inclusion of social preferences in the utility function to better explain the role of social concerns for economic behavior (Camerer & Loewenstein, 2004). This includes, for instance, preferences for fairness, trust and reciprocity, or altruism towards in-group members. However, while the role of social concerns has become well-established in various economic settings, their relevance is often ambiguous in the context of professional sports. For instance, formal and informal fairness norms are regarded as inherent features of sports competitions and are common knowledge for all agents involved. However, doping issues are a common phenomenon in many popular sports, questioning the relevance of fairness concerns. Moreover, diverse social and cultural backgrounds of key agents, such as athletes and teams, spectators, or sports referees and judges, may facilitate identity formation, leading to in-group favoritism. Such concerns about *fairness* and *group identity* become of even greater interest in the light of current global issues like political conflicts or the COVID-19 pandemic as well as recent internationalization processes in sports because such developments may lead to less trust in others' fairness or a stronger protection of in-group members.

From a sports economic and policy perspective, social concerns of spectators and judges are of particular interest because changes and deviations in sports demand and performance evaluations can have serious financial consequences for sports organizers

and competitors.[1] However, there is limited understanding about the role of social concerns for spectator and judge behavior for the following reasons:

First, fairness concerns and related trusting beliefs are generally regarded as important factors for economic behavior and demand (Kahneman et al., 1986). In professional sports, fairness is considered a precondition for spectator demand as it creates outcome uncertainty, making sports competitions attractive for spectators (Loland, 2001). However, fairness cannot be guaranteed because unethical and illegal behavior, such as match-fixing or doping, usually remains undetected and spectators need to trust in fair competitions (Dimant & Deutscher, 2019). In particular, doping is considered a serious issue since international superstars, sports organizations, and national institutions have been involved in past doping scandals (Gleaves & Hunt, 2016). In this regard, one line of research focuses on the impact of doping scandals on spectator demand. The existing empirical findings suggest that the doping–demand relation seems to be more complex than often expected and there is no evidence so far that trust in fairness works as the underlying key mechanism.

Second, identity concerns are also considered important for economic behavior and may lead to in-group favoritism. Identity formation often depends on the salience of social groups, like gender or nationality (Shayo, 2020). In professional sports, competitions are commonly organized by separating males and females and between athletes or teams from different countries. Thus, different social identities are salient in almost all sports-related contexts, especially at an international level. In this regard, one line of research focuses on national identity concerns and consumer behavior. Country of origin (COO) research suggests that political conflicts between importing and exporting countries increase the salience of consumers' national identity, affecting demand for brands and products from exporting countries in various settings (e.g., Pandya & Venkatesan, 2016). National identity also seems to be relevant for sports spectators; however, it is an open question whether countries' political relations also trigger spectators' national identity formation, and thus affecting the demand for sports. This issue is particularly important from a sports

[1] Chan et al. (2022) just recently emphasized the need to further explore social factors in professional sports settings to gain more insights into the behavior of the involved economic agents.

policy perspective because games of popular domestic cups and leagues are nowadays relocated to foreign countries in order to reach new target markets.

Third, another line of research focuses on national identity concerns of sports referees and judges. Unlike spectators, judges are professional experts who are paid to be impartial, meaning that their foul calls and performance evaluations should be unaffected by their nationality. Despite this, studies provide empirical evidence that referees and judges favor athletes of their own nationality (Pope & Pope, 2015; Sandberg, 2018). At the same time, however, studies also provide conflicting evidence on the potential sources and variation of this nationalistic bias. For instance, it was found to vary depending on career concerns of judges or the salience of group identity.[2] It therefore remains an open question whether this form of identity-based favoritism should be regarded as an inherent feature of judge behavior in international sports.

Based on the aforementioned issues, this dissertation aims to empirically explore the role of trust in fairness and national identity in the context of professional sports by focusing on spectator and judge behavior. The initial focus is on fairness and identity concerns on the demand side because spectators are considered the most relevant agents in professional sports (Borland & MacDonald, 2003). In fact, their steadily high interest has led to the professionalization and internationalization of the sports industry, making it a multi-billion-dollar business. Moreover, spectator behavior is unaffected by principal–agent relationships or career concerns, which allows the examination of their genuine preferences. Accordingly, Study 1 explores (i) the relation between trust in fairness and spectator demand and (ii) whether doping scandals reduce trust in fairness and consequently the demand for sports. Study 2 explores (iii) national identity concerns by testing the impact of countries' political relations on spectator demand.

While the first two studies suggest that national identity (rather than trust in fairness) is an important social concern for sports spectators, Study 3 aims to provide a more profound understanding of national identity concerns as an inherent feature of economic behavior in international sports. By focusing on the supply side of sports, it explores (iv) nationalistic bias in sports judges' performance evaluations and (v) its potential sources

[2] It is, for instance, generally argued that Olympic Games or world championships increase the salience of national identity compared to ordinary events because athletes represent their country and contribute to an overall country ranking (see Sandberg, 2018).

and variation. In this way, the empirical studies intend to contribute to the general and sports economics literature, with the aim to better understand the role of fairness and identity concerns for consumer and judge behavior in professional sports and beyond.

1.2 Structure of the dissertation

The remainder of this dissertation is structured as follows. Chapter 2 begins with putting the two concepts of trust in fairness and social identity into the research context of social preferences and economic behavior. It then continues with describing both concepts and the related empirical research and ends with a summary of the current state of literature, including research gaps and desiderata. Chapter 3 presents the research objectives that derive from the identified research gaps. Chapter 4 contains the three empirical studies conducted within the scope of this dissertation. Chapter 5 discusses their empirical findings and their theoretical and practical implications. Chapter 6 concludes and presents limitations and opportunities for future research.

2 Related literature

The two concepts of trust in fairness and national identity were recently established in behavioral economic research on social concerns and economic behavior. This chapter gives an overview of the research background and the related literature, including the relevant theoretical concepts and the current state of empirical research. Subchapter 2.1 describes the conceptual and empirical foundation of social concerns and puts the concepts of trust in fairness and social identity into context. Subchapters 2.2 and 2.3 present the related literature on trust in fairness and national identity, respectively. Subchapter 2.4 provides a summary of the state of literature and presents the identified research gaps and desiderata.

2.1 Social concerns and economic behavior

While initial work of behavioral economic research primarily focused on the role of risk and uncertainty for economic behavior (e.g., Kahneman & Tversky, 1979; Tversky & Kahneman, 1974),[3] the relevance of social concerns has subsequently emerged with theoretical and empirical contributions on social preferences and other-regarding behavior (Fehr & Schmidt, 2006). This line of research provided new and more profound explanations for behavior in various economic settings that involve social interaction. Models of social preferences and other-regarding behavior generally assume that people are not only motivated by their self-interest but are also concerned about "the payoffs allocated to other relevant reference agents and the intentions that led to this payoff profile in addition to the concern for one's own payoff" (Carpenter, 2010, p. 248).

Despite this all-encompassing definition, different conceptual models were established, focusing on and explaining behavior in different economic settings. The main social preference concepts were developed using experimental bargaining and cooperation games to reveal people's motives. Their findings raised concerns about the well-established self-interest assumption in economics. Most notably, findings from ultimatum games suggest that people not only care about *themselves* but also about the outcome of *others*, revealing fair behavior and punishment of unfair behavior when allocating money

[3] The initial ideas of behavioral economics can already be traced back to classical economics in the 18th century. For further details and information on the history and topics of behavioral economics see Camerer and Loewenstein (2004).

(Güth et al., 1982). Findings from related third-party punishment games suggest that people also care about outcomes *between other* reference agents due to fairness and cooperation norms (Fehr & Fischbacher, 2004). Likewise, findings from trust games suggest that people trust and believe in the fair behavior of reference agents, showing positive reciprocity instead of self-interested behavior (Berg et al., 1995; Cox, 2004). Findings from dictator games further suggest that not only fairness concerns explain behavior in ultimatum games but also altruistic preferences (Forsythe et al., 1994). Overall, these experiments provide empirical evidence of other-regarding behavior based on prosocial and antisocial preferences, such as fairness and equity, trust and reciprocity, or altruism and spitefulness (for a review and summary of the experimental games, see Fehr & Schmidt, 2006).

Based on these experimental findings, formal models were developed, explaining other-regarding behavior with social preferences, incorporating kindness and fairness concerns into people's utility function (e.g., Fehr & Schmidt, 1999; Rabin, 1993). In particular, the model by Fehr and Schmidt (1999) received much attention, making social preferences a popular topic in behavioral economics. According to their model, people are inequity-averse and willing to sacrifice some of their own material payoff to achieve more equitable outcomes in relation to others, explaining fair and (non-)cooperative behavior. Social preferences can also be *conditional*, depending on the reference agents' type or intentions. In this regard, Falk and Fischbacher (2006) combine inequity aversion with intention-based reciprocity, considering not just the material payoff but also the beliefs about the underlying fairness intentions. Charness and Rabin (2002) combine social-welfare preferences with reciprocity, considering the total payoffs of all agents and fairness intentions. Lastly, Levine (1998) describes interdependent preferences that are based on beliefs about other's altruism and spitefulness.

Overall, these well-established social-preference models refer to the idea that people compare themselves and others to other reference agents and / or consider the others' payoffs (Fehr & Schmidt, 2006). These models describe concerns for fairness, reciprocity, and altruism, which can be conditional or unconditional on the behavior and intentions of other reference agents. The formal models provide a good understanding on the nature and functioning of social preferences and emphasize the need to consider social concerns in economic behavior. They also describe the basic underlying principles for the

related concepts of trust in fairness and social identity, which are used in applied research to explore the impact of social concerns in real-life economic settings.[4]

Trust in fairness: Trust is considered an inherent feature of social life and economic activity (Knack & Keefer, 1997; Putnam, 1993). Furthermore, trusting beliefs and behavior are often based on the aforementioned fairness concerns. In particular, reciprocal and trusting behavior commonly reflect positive beliefs about the reference agents' fair behavior and intentions (Fehr, 2009). Again, findings from experimental games emphasized the role of trust in fairness. Güth et al. (1993, 1997) provided initial evidence, showing that it leads to cooperation and efficient outcomes in sequential ultimatum and trust games. Likewise, trust in positive reciprocity reflects trust in the fair behavior of reference agents in the trust game (Berg et al., 1995; Cox, 2004). In this regard, Kamas and Preston (2012) found that even different types of people (e.g., self-interested or inequity-averse) show high levels of trust in fairness. Similarly, Gächter et al. (2004) showed that trust in fairness also leads to more cooperation in public goods games. These experimental studies suggest that economic behavior can be conditional on trust in fairness. As such, fairness concerns are considered a key component of trust in many areas of economic activity.

Social identity: While the role of social preferences was soon recognized in behavioral economics, it remained an important question who the relevant reference agents are with whom people compare themselves and others. In everyday life, people do not care about all others equally. Identity concerns emerged as the most likely source and were incorporated into the aforementioned experimental games (e.g., Fershtman & Gneezy, 2001). Most notably, Chen and Li (2009) found that people are kinder to members of the same group, showing more charity, less envy, and positive reciprocity. Such in-group favoritism implies a form of altruism that is restricted to group members. Accordingly, identity-based preferences extend previous concepts of social preferences (Shayo, 2020). In principle, group identity can be artificially induced,[5] but in real life it usually relates to natural groups, such as gender or nationality. In this regard, Bernhard et al. (2006) and Goette et al. (2006) provide initial experimental evidence for in-group favoritism toward

[4] The formal models and preference types are not easily applicable to many economic phenomena as discussed by Fehr and Schmidt (2006).

[5] Artificially induced group identity based on random or preference-based assignment to meaningless groups is frequently used in experiments and known as the minimal group paradigm (Chen & Li, 2009).

reference agents with the same ethnicity and from the same platoon, respectively.[6] Overall, these experimental findings suggest that identity concerns affect economic behavior. This issue seems particularly relevant because people are members of many social groups, which are salient in many areas of economic activity.

Trust in fairness and social identity can thus be considered extended concepts of social preferences that are conditional on other reference agents. Moreover, the concepts may help to explore the role of fairness and identity concerns for behavior in real-world economic settings *in general* and professional sports *in particular*, as described in detail in the following Subchapters 2.2 and 2.3.

2.2 Trust in fairness and economic behavior

While the role of trust in fairness for economic behavior was initially shown in experimental studies on social preferences and cooperative behavior, it seems particularly relevant in settings where fairness concerns are based on existing norms.[7] One such setting is professional sports, where fairness is considered an inherent feature of sports competitions. It relates to formal and informal fairness norms, such as fair play and mutual respect, which ensure compliance with the rules and equal opportunities for the competing athletes. These fairness norms are common knowledge and used to maintain the integrity of sports competitions. Accordingly, fairness is considered a central element of sports, encouraging athletes to enter competitions and used to justify public funding for sports in many countries.[8]

In professional sports, the relevance of fairness concerns is predominantly motivated by the argument that spectators' consumption preferences are based on natural top-level athletic performances and fairness of the competition. Both features are consumed simultaneously and expected to create suspense and outcome uncertainty (Buechel et al., 2016; Loland, 2001), making sports interesting and attractive for spectators (Neale, 1964; Rottenberg, 1956).[9] It is thus expected that athletes, sports organizers, spectators, and the

[6] Goette et al. (2006) also found evidence for out-group hostility.
[7] For a general discussion on community standards of fairness and behavior in markets see Kahneman et al. (1986).
[8] For instance, see the policy program of the United Kingdom (H.M. Government, 2015).
[9] For a recent summary and discussion on the relevance of suspense and outcome uncertainty in sports see Pawlowski and Nalbantis (2019) and Pawlowski et al. (2018).

general public prefer fair competitions (Bird & Wagner, 1997). In contrast, however, unfair behavior, such as doping and match-fixing, is a common phenomenon.[10] More importantly, fairness cannot be guaranteed because such unethical and illegal behavior usually remains undetected or is revealed long after the sports competitions took place. All agents involved therefore need to trust in a fair competition. This particularly applies to spectators because they are the least able to monitor unethical or illegal conduct.[11]

In this context, one line of research focuses on the impact of doping scandals on spectator demand. Doping is widely seen as a threat to sports worldwide and society at large. Even the European Commission (2007) and the UNESCO (2005) – organizations at the highest political level – adopted sports political actions against doping. Similarly, many countries established anti-doping laws to protect the social values and integrity of sports. This issue seems particularly delicate since international superstars and sports organizations as well as national institutions have been involved in past doping scandals.[12] It is thus a serious concern and popular claim that doping scandals reduce spectator sports demand, driven by a loss of trust in fairness (Dimant & Deutscher, 2019; Frenger et al., 2013). For instance, the public television (TV) channels in Germany stopped broadcasting the Tour de France in 2012 for three years after a considerable decline in the audience interest. Doping issues were stated as the main reason, although the absence of national cycling superstars may also have led to the drop in TV ratings (Van Reeth, 2013). Understanding the impact of trust in fairness on spectator demand and its role in the doping–demand relation is therefore a relevant issue from an economic and sports policy perspective.

What follows is a brief presentation of the trust in fairness concept (Section 2.2.1) and the current state of empirical research on trust in fairness and demand-related behavior, specifically focusing on doping scandals and spectator behavior (Section 2.2.2).

2.2.1 Concept of trust in fairness

Fairness concerns are by now regarded as a strong motive for human behavior and often impact social, political, and economic measures. They generally describe people's

[10] For reasons why athletes engage in doping abuse see Bird and Wagner (1997) or Haugen (2004).
[11] A similar line of argumentation is used for other consumer product categories, such as handicrafts or food, where fairness concerns relate to fair working conditions or similar issues. As in professional sports, consumers cannot easily monitor fairness and need to trust, for example, fair trade labels.
[12] For further details and information on the history of doping in sports see Gleaves and Hunt (2016).

willingness to "sacrifice personal gains in order to eliminate inequalities they view as unfair" (Almås et al., 2010, p. 1176). Furthermore, fairness concerns relate to a person's perception of distributive justice (i.e., fair outcomes) as well as procedural justice (i.e., fair processes), which both appeal to common moral norms that compete with self-interest in a given situation. These norms contain principles of equality, utilitarianism, or equity, and depend on the context in which fairness is evaluated (for a detailed description, see Konow, 2003). In sports, what is considered fair is subject to change but generally refers to principles and rules that ensure competitors' responsibility for their performances (e.g., ability and training effort), comparability of performances, as well as mutually respectful and virtuous behavior (Loland, 2001).

From a consumer demand perspective, fairness is often accompanied by "the element of trust" (Kahneman et al., 1986, p. 736). Trust can be defined as the "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al., 1995, p. 712). This is important in almost all types of social and economic relations;[13] however, what constitutes trust has been a matter of debate (for a discussion, see Fehr, 2009). In particular, findings from the aforementioned trust game of Berg et al. (1995) led to the development of numerous trust models, describing cognitive and behavioral concepts of trust.

The concept of spectators' trust in fairness is based on the commonly used and generally accepted two-component trust concept by McKnight et al. (1998). This concept distinguishes between the trusting intention, which is the willingness to depend on another person, and trusting beliefs in the other's ability, benevolence, and integrity. In this concept, the trusting intention and integrity beliefs often relate to an individual's fairness concerns. In particular, integrity describes trust in the moral and ethical conduct of the reference agent, and, as such, represents fairness beliefs in the context of spectator sports demand.[14] In this regard, top athletes can be considered the most relevant referent

[13] For instance, previous research related trust to economic growth (Knack & Keefer, 1997), international trade (Guiso et al., 2009), or investment decisions (Guiso et al., 2008).
[14] Trusting beliefs of benevolence may also capture fairness concerns in different settings but are less relevant in relation to sports.

agents, acting as trust or distrust ambassadors who signal the fairness or unfairness of the competition to the spectators, respectively.

2.2.2 Empirical research on trust in fairness and spectator behavior

Besides the experimental evidence that trust in fairness is a precondition for cooperative behavior (see Subchapter 2.1), only few studies focused on its role for demand-related behavior. In their work on fairness in market transactions, Kahneman et al. (1986) provide initial and descriptive evidence that fairness concerns can motivate consumer behavior. Moreover, Guiso et al. (2008) used micro- and cross-country survey data to show that trust leads to more and higher investments in the stock market, explaining this finding with the need to trust the fairness of the financial system. Similarly, Giannetti and Wang (2016) found that fraud scandals of firms decrease households' investments in stocks due to a loss of trust in the stock market. In the context of food markets, Ding et al. (2013) also found that trust mitigated consumer reactions after food scandals.

Although the aforementioned studies emphasize the impact of scandals on trust and the role of trust in fairness for demand-related behavior, comprehensive evidence does not yet exist. Likewise, neither the impact of doping scandals on trust in fairness nor its impact on spectator behavior and demand are well established in the sports economics literature. To the best of my knowledge, only Buraimo et al. (2016) showed that a major corruption scandal decreased home-game attendance in Italian soccer. However, their attempt to empirically explain the underlying mechanism does not support moral values as the driver in this relation. Using a qualitative approach, Manoli et al. (2020) also argued that people do not even trust or believe in the integrity of sports but are still interested in watching sports. Both studies focus on similar concepts like trust in fairness but cannot explicitly identify the role of fairness concerns for spectator demand.

Furthermore, empirical studies on the effect of doping scandals on spectator demand showed that doping suspensions in Major League Baseball (MLB) decreased home-game attendance (Cisyk & Courty, 2017) and the TV audience size of the affected teams (Cisyk, 2020). These effects lasted for about two and five weeks, respectively. Negative spillover effects on other MLB teams are found for stadium attendance but not for TV demand. While the authors could not directly identify the underlying mechanism, the mere absence of superstar athletes was ruled out as an explanation for the negative doping effects in

MLB (Cisyk, 2020; Cisyk & Courty, 2017).[15] In contrast to the previous findings, Brave and Roberts (2019) found no impact of doping suspensions on MLB's gate revenues per season. Studies focusing on the TV audience in cycling found inconsistent short-term and strong long-term effects of doping scandals on TV demand in Flanders, Belgium (Van Reeth, 2013), but no short-term effects in Denmark (Feddersen, 2020), and no long-term effects in Spain (Rodríguez et al., 2015). To explain his inconsistent findings, Van Reeth (2013) argues that spectators may also derive utility from witnessing scandals and related media coverage. Scandals may thus be part of the overall sports entertainment product and not necessarily reduce spectator demand. In contrast, however, an experimental study by Abeza et al. (2020) found that fans of the Australian Football League view doping as a serious offense that negatively affects their viewing intentions.

Overall, no study has yet directly tested the impact of doping cases or comparable scandals on spectators' trust in fairness. Some studies indicate that doping scandals can reduce spectator demand, but this finding is inconsistent within and between the different sports. The doping–demand relation thus seems to be rather complex and there is limited understanding of the underlying mechanism(s).

2.3 National identity and economic behavior

While social identities are regarded as a likely source for social preferences and other-regarding behavior in many real-life settings, people may identify with different social groups in a given situation. These groups can be permanent or temporary and induced or natural. Most notably, economists have a serious interest in understanding identity concerns relating to salient natural groups because they can affect behavior in many areas of economic activity. This includes, for instance, the role of religious or national identities in trade relations and political conflicts as well as gender, racial, or ethnic favoritism and discrimination in the labor market.[16] In professional sports, different social identities are salient in basically all contexts. National identity seems to be particularly important

[15] Several studies show that superstars with exceptional skills drive spectator demand for several sports (e.g., Brandes et al., 2008; Hausman & Leonard, 1997). Accordingly and in line with the economics of superstars (MacDonald, 1988; Rosen, 1981), reduced demand after doping scandals could thus also be driven by lower product quality due to the absence of banned superstar athletes rather than a loss of trust in fairness.

[16] For a recent collection of empirical evidence on the role of social identity in various economic contexts see Shayo (2020).

because top-level competitions commonly reach a global audience and are often organized at an international level. This leads to diverse cultural and political backgrounds of spectators, athletes, and judges, etc.

In this regard, one line of research focuses on countries' political relations and spectator sports demand. This is regarded as particularly relevant for popular domestic competitions, such as the National Football League (NFL) or the Spanish super cup in soccer, who have strengthened their internationalization efforts in recent years to reach foreign target markets. Their internationalization strategies include, for instance, deals with television and streaming platforms to increase broadcasting time, preseason marketing trips to the target markets, or the relocation of games. Most notably, the latter strategy has become increasingly important in recent years in the popular North American major leagues and European soccer leagues. In this context, spectators' positive (or negative) perception of the political relations between importing country and COO of the sport is expected to make similarities (or differences) between both countries salient and mitigate (or facilitate) national identity formation. Understanding the impact of countries' political relations on spectator behavior may thus help to shed light on the role of national identity concerns for spectator sports demand.

Another line of research focuses on national identity concerns of sports referees and judges. Unlike spectators, they are professional experts in their task and paid to be impartial. Their foul calls and performance evaluations should thus be unaffected by their social identities. However, in international sports competitions, national identity concerns might affect their decisions, which may lead to in-group favoritism toward compatriot athletes. In the worst case, this nationalistic bias can affect contest results, the distribution of prize money, and, accordingly, the sporting and economic success of athletes. Understanding whether nationalistic bias is an inherent feature of judge behavior in professional sports is therefore of considerable interest from a sport economic and policy perspective. Research on this matter may also shed more light on potential sources and drivers of in-group favoritism bias in international sports and other settings that involve subjective evaluations of experts (e.g., student evaluations at international schools and universities, hiring and promotion processes of global firms and organizations, judicial sentencing, or policymaking decisions).

What follows is a presentation of the social identity concept with a particular focus on the role of nationality (Section 2.3.1). Subsequently, the current state of empirical research on national identity and spectator behavior (Section 2.3.2) and judge behavior (Section 2.3.3) is presented.

2.3.1 Concept of social (and national) identity

Like fairness concerns, identity concerns are regarded as a strong motive for human behavior. Most notably, social psychologists Tajfel and Turner (1979, 1986) developed social identity theory to explain intergroup behavior, such as violent conflicts and competition. According to them, social identity consists "of those aspects of an individual's self-image that derive from the social categories to which he perceives himself as belonging" (Tajfel & Turner, 1986, p. 283). In this way, social groups provide a person with a sense of identity and belonging, maintaining or enhancing one's self-esteem and self-concept. Here, a person distinguishes between the own in-group and a relevant out-group to achieve a favorable comparison. This is based on a mental process, in which a person first categorizes the own social environment, then identifies with a social group, and finally compares the own in-group with other groups (Tajfel & Turner, 1986).

While social identity theory was developed to explain intergroup behavior with serious economic consequences, it was Akerlof and Kranton who formally introduced the identity concept into economics in 2000. In their seminal paper, they incorporated social identity into a general utility function, in which identity and related social norms affect preferences and motivate behavior. In this regard, agents' payoffs are based on their own identity-based behavior and the identity-based behavior of others. Accordingly, social identity is expected to impact organizational and consumer behavior as well as behavior in the education sector or in the labor market (Akerlof & Kranton, 2000).

Nonetheless, their contribution leaves the question open as to which social group people identify themselves with in real economic situations. People usually belong to many different groups that are considered as more or less important. For this reason, Shayo (2020) further specifies the identity-based preference model of Akerlof and Kranton (2000), adding that social identity formation depends on the group status and one's perceived distance from the group attributes. Group status reflects the individual's value

of belonging to a given group and is based on comparisons to other groups. The perceived distance between an individual and a social group is based on the own attributes and the prototypical salient attributes of the group members. Utility maximization can then be achieved by increasing the status of one's own group, by reducing one's perceived difference to that group, or by changing the group one identifies with.

According to Shayo (2020), this concept of social identity implies that identity formation is more likely if groups have a higher status and if a given social group is perceived as more similar in its salient attributes. Utility is then maximized by showing favoritism towards in-group members. The model thus extends initial social preference concepts because in-group favoritism is understood as altruistic behavior to increase group status. In addition, people can identify with multiple groups to a variable extent, depending on the status and salience of group attributes in a given situation. Choosing one's identity may thus vary by social context.

Generally speaking, the role of social identity types for economic behavior can be ambiguous and a matter of debate because the status and salience of group attributes is often unclear. However, national identity concerns seem highly prevalent, especially in professional sports. This is because national attributes are relatively stable and well-known, including, for instance, the same language, food culture, explicit symbols (national flags, emblems, anthems), and a large set of social norms. The salience of national attributes can also easily be increased through various means, like the use of national symbols, but also by environmental factors, like political conflicts or threat of war. Moreover, nationality also has a relatively high status compared to alternative group identities like social class (Shayo, 2009).

2.3.2 Empirical research on national identity and spectator behavior

Besides the evidence that identity concerns affect behavior in experimental games (see Subchapter 2.1), a large body of literature focuses on COO effects on consumer behavior. Research on this topic indicates that COO information affects product evaluations and purchase decisions because consumers use COO cues to infer product quality but also to associate a product with their national identity (for a review, see Verlegh & Steenkamp, 1999). A subarea of COO research particularly focuses on countries' political relations and the role of consumer animosity towards foreign countries. Klein et al. (1998) were

the first to show a negative impact of consumer animosity on purchase decisions, which was unrelated to product quality perceptions, suggesting that national identity affects consumer preferences and demand. Further studies confirmed this initial finding for different importing countries (where the study was conducted) and exporting countries (the products' COO), product types (e.g., cars or electronic devices), and sources of animosity development (e.g., political, economic, or historical reasons; for a review, see Riefler & Diamantopoulos, 2007).

The most conclusive evidence for national identity concerns stems from studies on international political conflicts and subsequent consumer boycott calls. In the context of the Iraq war in 2003, Chavis and Leslie (2009) found a decrease in US sales of French wine; Pandya and Venkatesan (2016) found a decrease in the market share of French-sounding brands in US supermarkets; and Clerides et al. (2015) found a decrease in sales of US soft drinks in seven Arab countries. In particular, Pandya and Venkatesan (2016) provide empirical evidence that national identity is the most likely mechanism that drives consumers' boycott behavior. Likewise, Heilmann (2016) analyzed the impact of several international political conflicts on consumer, intermediate, and capital goods, finding that boycott calls were most effective for consumer goods with a salient COO-branding.

Overall, these findings suggest that international political conflicts increase the salience of national identity, affecting consumer demand. However, the aforementioned studies explored consumer animosity and boycott calls in the context of historic or current crises between two countries, but little is known about the role of political relations in the absence of severe conflicts. Moreover, the studies typically focused on durable goods, such as cars or electronic devices, but consumers who want to avoid brands from less favorable countries are expected to sacrifice less if products are non-durable and easy to substitute (Riefler & Diamantopoulos, 2007). The latter applies in particular to entertainment products like sports events, suggesting that COO effects may also shape spectator demand.

So far, however, the impact of countries' political relations on spectator behavior and demand has not yet been explored. Two studies analyzed spectator demand for transnational TV broadcasts but focused on game uncertainty (Nalbantis & Pawlowski, 2019; Schreyer et al., 2018). Moreover, most studies exploring the role of spectators'

social identities focused on fans' identification with their favorite teams (e.g., Laverie & Arnett, 2000; Shtudiner et al., 2021). Only few studies have explored the role of spectators' national identity but only for the TV demand of international competitions. Bennett et al. (2007) showed that patriotism was positively associated with following the English cricket team. Nüesch and Franck (2009) showed that spectators' national identity increases the TV demand for soccer games of the own national team in Switzerland – a country with a large share of residents with a foreign nationality. Chiang and Jane (2013) also found higher TV demand for games of the World Baseball Classic in Taiwan if the national baseball team was playing. An increase in TV audience ratings was also found if well-performing compatriot athletes participate in cycling races (Van Reeth, 2013) and tennis matches (Konjer et al., 2017), while this was not found in Formula One motor racing (Schreyer & Torgler, 2018).

Overall, the COO literature suggests that consumer animosity toward foreign countries and boycott calls negatively impact consumer demand. Likewise, studies emphasize that national identity concerns affect spectator demand for most sports played at an international level; however, no study has yet tested the impact of countries' political relations on spectator sports demand.

2.3.3 Empirical research on national identity and judge behavior

Besides the aforementioned evidence that national identity affects consumer behavior, it has also been related to judging bias of professional experts. These experts are appointed by principals if high quality judgements are needed and expected to be impartial. However, experts' national identity is a likely source of in-group favoritism that leads to judging bias in international evaluation settings.

Surprisingly, the literature on nationalistic bias and subjective evaluations in *international settings* outside professional sports is scarce. To the best of my knowledge, only one study by Feld et al. (2016) found evidence for nationalistic bias among graders of university exams. This insufficient number of studies is likely due to a limited quality and quantity of available data. However, a large body of literature focuses on similar judging biases in *national settings* that also relate to differences in origin, i.e., ethnic and racial in-group favoritism. These judging biases were consistently found for evaluations in education and training (Bar & Zussman, 2020; Dee, 2005), hiring and promotion

processes (Åslund et al., 2014; Giuliano et al., 2009, 2011), and in judicial sentencing (Anwar et al., 2019; Gazal-Ayal & Sulitzeanu-Kenan, 2010; Grossman et al., 2016; Lim et al., 2016; Shayo & Zussman, 2011).[17] Shayo and Zussman (2011) also found that ethnic bias of legal judges varies with the salience of ethnic identity. Moreover, there is evidence for ethnic and racial bias among sports referees (Mongeon & Longley, 2015; Parsons et al., 2011; Pope et al., 2018; Price & Wolfers, 2010). Overall, these findings suggest that economic agents commonly use their origin for social identity formation, which is a major driver of judging bias in national evaluation settings.[18]

Furthermore, there is consistent evidence for the impact of national identity concerns on judging behavior in international sports. Nationalistic bias was found in referee decisions in soccer (Pope & Pope, 2015) and rugby (Page & Page, 2010), as well as in performance evaluations in diving (Emerson et al., 2009), dressage (Sandberg, 2018), figure skating (Zitzewitz, 2006, 2014), and ski jumping (Lyngstad et al., 2020; Scholten et al., 2020; Zitzewitz, 2006). The most profound insights stem from analyses of nationalistic bias variation. For instance, Scholten et al. (2020) suggest that nationalistic bias increases with judges' age. Most notably, Sandberg (2018) suggests that nationalistic bias of compatriot judges and their panel members is mainly driven by the salience of (temporary) national identity, and thus resembles findings from Shayo and Zussman (2011) on ethnic bias. However, strategic concerns and incentives for consistent judging of panel members may also explain her findings, at least to a certain extent. A similar claim was made by Zitzewitz (2006), who showed that judges vary their nationalistic bias strategically according to the competition stakes and their own career concerns, and that judges also engage in vote trading. Moreover, Pope et al. (2018) replicated the study by Price and Wolfers (2010) on referees' racial bias, showing that the bias disappeared after broad media attention. Therefore, findings from studies using sports settings suggest that the

[17] Depew et al. (2017) also found a negative racial in-group bias in juvenile courts, which was explained with the defendants' violations of social group norms.

[18] From a social and economic policy perspective, another important social identity relates to gender. However, empirical evidence on gender in-group bias is mixed. Studies report same-, opposite-, and predominately no gender favoritism in education and training (Bar & Zussman, 2020; Dee, 2005; Feld et al., 2016), hiring and promotion processes (Bagues & Esteve-Volart; 2010; Bagues et al., 2017; De Paola & Scoppa, 2015), and judicial sentencing (Boyd et al., 2010; Gruhl et al., 1981; Hoekstra & Street, 2021; Knepper, 2018; Lim et al., 2016). This can be (partly) explained by the relatively lower salience of gender identity in the given situation and an arguably lower group status.

impact of in-group favoritism seems quite volatile. It is therefore still unclear whether nationalistic bias should be regarded as an inherent feature of judge behavior.

2.4 Summary and research desiderata

The initial work on social preferences suggests that social concerns play an important role for economic behavior in different areas of economic activity. From a sports economic and policy perspective, fairness and identity concerns are considered particularly relevant, and there is supporting literature suggesting that these social concerns may matter for spectator and judge behavior in professional sports. However, several key issues have not yet been addressed. Based on the review of the current state of research, overall, five research gaps and desiderata are identified and discussed in the following. The first three relate to fairness and identity concerns and economic behavior on the demand side of professional sports, focusing on spectator behavior; the latter two relate to identity concerns on the supply side, focusing on judge behavior.

First, the general economics literature suggests that scandals of unethical and illegal conduct may lead to a loss of trust in fairness and consequently reduce consumer demand. This is also considered a serious issue in professional sports because fairness is an inherent feature of sports competitions. While theoretical contributions emphasize the importance of trust in fairness for spectator demand, no empirical study has yet empirically tested this relation. Whether spectators care about fairness in sports competitions can thus neither be confirmed nor ruled out.

Second, trust in fairness seems to be particularly relevant in international sports with a dark history of doping abuse. However, no study has yet tested the impact of doping scandals on spectators' trust in fairness. Moreover, the few existing studies on the short- and long-term effects of doping scandals on spectator sports demand report inconsistent findings within and between different sports, suggesting that the doping–demand relation is more complex than often assumed or claimed. The literature on the doping–demand relation is therefore still inconclusive and lacks a direct test of the potential key mechanism, i.e., trust in fairness.

Third, recent internationalization efforts of popular domestic sports emphasize the need to explore the impact of countries' political relations on sports demand to better

understand spectators' identity concerns when sports are considered *imported* goods. Literature on COO effects suggests that political conflicts between importing and exporting countries increase the salience of national identity, resulting in animosity and boycott calls that can have a substantial impact on consumer behavior and demand. However, it is less clear whether the perceived status of countries' political relations also increases the salience of consumers' national identity when conflicts are absent. This issue seems particularly relevant for products that are non-durable and easy to substitute – like spectator sports. In this regard, a few studies showed that national identity also matters for spectator demand but empirical evidence on the impact of countries' political relations on demand for professional sports *in general* and popular domestic competitions *in particular* is so far missing.

Fourth, while spectators are generally expected to state and reveal their genuine preferences, one expects that professional agents should be able to resist their inherent preferences in situations where they are paid to act in the interest of their principals. Therefore, exploring nationalistic bias of judging experts allows for a more profound understanding of the relevance of identity concerns for economic behavior. While the related literature found ethnic and racial biases in *national* evaluation settings, several studies also provide empirical evidence for nationalistic bias among referees and judges in *international* sports. However, it remains an open question whether nationalistic bias should be regarded as an inherent feature and prevalent issue in sport performance evaluations or whether national identity concerns only affect a small share of judges.

Fifth, previous studies on nationalistic bias of sports judges provide conflicting evidence on its potential sources. Findings from dressage competitions suggest that it may be driven by the salience of national attributes and identity-based preferences. Findings from figure skating and ski jumping competitions suggest that it may rather be driven by strategic concerns. Besides, nationalistic bias may also be reinforced or mitigated by the judging systems employed by the different sports. It therefore remains unclear whether variation of nationalistic bias rather reflects the salience of social groups and the inherent preference toward in-group members, or whether it is driven by strategic concerns and economic incentives. Accordingly, a better understanding of its potential sources and variation is much needed.

3 Research objectives

This chapter describes the research objectives and empirical approaches of the studies conducted within the scope of this dissertation (presented in Chapter 4). The three studies address the previously presented research gaps and desiderata to provide new and further insights into the relation between social concerns and economic behavior in professional sports. For this purpose, they focus on sports settings where fairness and national identity concerns are considered particularly relevant, i.e., spectators' TV demand, spectators' on-site demand, and judges' performance evaluations. In this way, the studies aim to contribute to the sports economics literature and inform economic research beyond professional sports.

The initial focus is on fairness and national identity concerns of spectators because understanding sports consumer behavior and demand is regarded as the most relevant issue of economic research on professional sports (Borland & MacDonald, 2003). The latter focus is on national identity concerns of sports judges, which may lead to in-group favoritism toward compatriot athletes. To explore each of the settings and to test the relations of interest, Study 1 and Study 2 draw on primary data from individual-level panel surveys because testing the role of consumers' fairness and identity concerns with higher aggregated secondary data appeared to be difficult in previous research. Study 3 draws on secondary data of sports performance evaluations at the individual judge level because it allows for a clean identification strategy to reveal nationalistic bias. The data of the three studies therefore allow explicit and direct tests of the relations of interest, as described in the following.

Study 1: Otto, F., Pawlowski, T., & Utz, S. (2021). Trust in fairness, doping, and the demand for sports: a study on international track and field events. *European Sport Management Quarterly*, *21*(5), 731–747.

Study 1 is presented in Subchapter 4.1 and focuses on fairness concerns of sports spectators. The study uses data from a two-wave panel survey and a sample that is representative for the adult German population and consists of two parts, addressing the first and second research gap, respectively. The first part focuses on the full panel survey

to test the relation between trust in fairness and the revealed TV consumption of an international track and field event by employing regression estimations. The second part focuses on a real doping case and a sports event scenario that was presented to a randomly drawn subsample in the second survey wave. Here, stated consumer preferences (willingness-to-pay (WTP) and intention-to-consume (ITC)) were elicited to proxy spectators' TV demand. Radius matching estimations based on the propensity score were employed to test the relation between the doping scandal, trust in fairness, and TV demand. In this way, the study attempts to contribute to the scarce literature on trust in fairness and spectator demand. Moreover, it aims to provide new insights into the doping–demand relation, focusing on athletes as trust ambassadors and international track and field events – a sport where doping scandals are a common phenomenon. As such, it also contributes to the literature on doping and spectator demand, with the aim to explain some of the previous and inconsistent findings. It also aims to provide further insights on fairness concerns and general consumer behavior, using a setting where fairness is a central product feature. The study also considers the potential impact of patriotism – a measure of national identity concerns – on spectator demand for an international competition, and thus encourages further research into this issue, as is done in Study 2.

Study 2: Otto, F., Nalbantis, G., & Pawlowski, T. (2022). Political relations and sports: Exploring the demand for relocated soccer games. *European Sport Management Quarterly*, 1–19.

Study 2 is presented in Subchapter 4.2 and focuses on countries' political relations and spectator demand. Addressing the third research gap, the study uses two online surveys and a sample that is representative for soccer-interested US residents to elicit the stated preferences (i.e., WTP) for relocated cup finals and league games of European soccer leagues in a hypothetical scenario. Tobit regression and instrumental variable (IV) estimations are employed to explore the relation between the perceived friendliness between the countries of origin and the US as the importing country and the WTP for admission to the relocated games. While, theoretically, countries' political relations trigger national identity formation, people are also expected to vary in their initial level of national identity. This may affect the attitude toward foreign countries in general and

make individual's identity formation more likely. The IV approach exploits this and uses an instrument that proxies the level of national identity in the respondents' area of residence. The main objective of this study is to provide first insights into the relation between countries' political relations and spectator sports demand, with a focus on relocated games – a common internationalization strategy of popular domestic sports leagues. In this regard, it attempts to contribute to the literature on COO effects in sports and spectators' national identity concerns. It also aims to contribute to the general COO literature by focusing on a product that is non-durable and easy to substitute and a consumption setting where severe political conflicts are currently absent.

Study 3: Krumer, A., Otto, F., & Pawlowski, T. (2022). Nationalistic bias among international experts: Evidence from professional ski jumping. *The Scandinavian Journal of Economics, 124*(1), 278–300.

Study 3 is presented in Subchapter 4.3 and focuses on national identity concerns and economic behavior on the supply side of sports, i.e., nationalistic bias of sports judges. The study uses a large dataset of performance evaluation scores from international ski jumping competitions of seven seasons to compare scores of compatriot judges with the scores of their panel members, addressing the fourth and fifth research gap. Multiple-level fixed effects regressions are employed to identify nationalistic bias in judges' performance evaluations and its potential sources and variation. In this regard, the study attempts to replicate previous findings on nationalistic bias in ski jumping with up-to-date data to see whether this issue is still prevalent in performance evaluations or whether most judges are able to resist their preferences toward compatriot athletes. After that, the study extends and compares previous analyses on nationalistic bias of sports judges to further understand potential sources and variation. In this way, the study contributes to the literature on nationalistic bias of judges in international sports. It further aims to contribute to the general literature on replication studies and in-group favoritism of professional experts. In this regard, evaluations in professional sports resemble other settings where subjective judgments have serious economic consequences and where stakes are high, making it ideal for exploring the role of national identity concerns in experts' decision-making.

4 Empirical studies

This chapter contains the three empirical studies conducted within the scope of this dissertation. All studies follow the style guidelines of the journal where the study is published or currently under review (including paper structure, citation style, wording, etc.). Study 1 is published in a special issue on the societal impact of elite sports in the *European Sport Management Quarterly* (for information on the special issue, see De Bosscher et al., 2021) and uses the wording suggested by the call for papers (e.g., "elite sports" instead of "professional sports"). Study 2 is also published in the *European Sport Management Quarterly* and Study 3 is published in *The Scandinavian Journal of Economics.* All studies include an introduction and further sections on the data and sample, variables and empirical strategy, results and discussion, and a conclusion. At the end, each study includes a reference list and an appendix with further information on the data and results (e.g., robustness checks). In case the appendix for the published paper is provided as an online document, and thus labeled as "Online Appendix" in the study, it is further specified as "Appendix (online appendix)" in this dissertation.

4.1 Trust in fairness, doping, and the demand for sports: A study on international track and field events*

EUROPEAN SPORT MANAGEMENT QUARTERLY
2021, VOL. 21, NO. 5, 731–747
https://doi.org/10.1080/16184742.2021.1942125

Trust in fairness, doping, and the demand for sports: a study on international track and field events

Felix Otto [a,b], Tim Pawlowski [a,b,c] and Sonja Utz [d,e]

[a]Faculty of Economics and Social Science, Institute of Sports Science, University of Tübingen, Tübingen, Germany; [b]Interfaculty Research Institute for Sports and Physical Activity, University of Tübingen, Tübingen, Germany; [c]LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany; [d]Social Media Lab, Leibniz-Institut für Wissensmedien, Tübingen, Germany; [e]University of Tübingen, Tübingen, Germany

ABSTRACT

Research question: We investigate empirically the impact of spectators' trust in fairness of the sporting competition on TV demand for sports as well as its role in the doping-demand relation.
Research methods: We use data from a two-wave panel survey and a sporting event scenario where doping issues are present. The effects are estimated using Zero-Inflated Poisson regressions and radius matching based on the propensity score.
Results and findings: Results of the panel survey analysis suggest that trust in fairness has no impact on TV demand. Our scenario analysis further reveals that awareness about a major doping case negatively affects trust in the fair conduct and integrity of athletes. However, we again find no general effect on the demand for sports.
Implications: Results of our study suggest that athletes can be seen as (dis)trust ambassadors promoting the (un)fairness of the sporting competition to consumers. However, in contrast to popular (political) claims, our results question the trust channel as a driver for spectator sports demand and as a mechanism in the doping-demand relation.

ARTICLE HISTORY
Received 3 August 2020
Accepted 7 June 2021

KEYWORDS
Trust in fairness; trust ambassador; doping; spectator sports demand; track and field

*Acknowledgements: A previous version of the paper was presented at the 4th Sport Economics & Sport Management (SESM) Conference in Berlin, Germany, 2019; the 11th Conference of the European Sport Economics Association (ESEA) in Gijon, Spain, 2019; and the 24th University Day of the German Society of Sport Science (dvs) in Berlin, Germany. We thank the participants for their valuable comments and suggestions. The authors would also like to thank Antonia Heppeler. The usual disclaimer applies.
Funding: This work was supported by joint funding from the Institute of Sports Science, University of Tübingen and the Leibniz-Institut für Wissensmedien.

Trust in fairness, doping, and the demand for sports: A study on international track and field events

Introduction

Trust is a core dimension of social capital and an inherent feature of social life (Putnam, 1993). As such, trust is considered an important factor in many political and economic situations (e.g. Guiso et al., 2008; Hetherington & Husser, 2012). In particular, a loss of trust due to scandals and unethical or illegal conduct by agents is frequently associated with reduced support and demand. For instance, Chanley et al. (2000) found that political scandals decreased public trust and support for policy measures. Ding et al. (2013) found that trust mitigated negative consumer reactions to bovine spongiform encephalopathy (BSE) incidents in Canada.

Similar concerns are often discussed in context of spectator sports, where different kinds of corruption and manipulation have been observed since ancient times (Maennig, 2005). This includes match-fixing, the misuse of insider information, and the abuse of performance-enhancing drugs (PED), i.e. doping. In particular, the most recent doping scandals in sports such as track and field (T&F) and cycling have caused substantial debates and (sport) political actions,[1] since doping is seen as a threat to sports worldwide and society at large (European Commission, 2007; United Nations Educational, Scientific and Cultural Organization, 2013). In this regard, trust in fairness is often seen as a precondition for spectator sports demand and a potential mechanism after scandals of unethical or illegal conduct. The assumption is that consumer preferences for watching elite sports are inherently based on fairness of the sporting competition. However, athletes' unethical and illegal conduct such as doping is usually hidden and sport spectators therefore need to trust that athletes will compete fairly. In case of a scandal, consumers might lose their trust in fairness, which in turn leads to a decline in sports demand (Preston & Szymanski, 2003).

[1] For instance, Germany established an anti-doping law in 2015 (Federal Ministry of Justice and Consumer Protection, 2015). Moreover, the last amendments of the World Anti-Doping Code came into force in 2019 (World Anti-Doping Agency, 2019).

Theoretically, this causal chain sounds plausible because top athletes may act as (dis)trust ambassadors signaling the (un)fairness of the entire field of athletes (McEvily et al., 2003). However, no study has yet directly investigated the role of trust in fairness for spectator sports demand. Therefore, the objective of this paper is twofold: First, we examine the role of trust in fairness for spectator sports demand *in general*. Second, we investigate whether a doping scandal can cause a loss of trust followed by reduced demand *in particular*. Our setting is T&F, which has been affected by several doping scandals in the past. The data come from a representative panel sample of adults living in Germany surveyed online in two consecutive waves. In the first part of our analysis, we regress the participants' TV consumption of a major T&F event on a trust in fairness measure. In the second part of our analysis, we make use of a sporting event scenario and follow a selection-on-observable approach with radius matching estimations. Overall, while awareness about a major doping case negatively affects trust in the fair conduct and integrity of athletes, this in turn does not affect the demand for sports.

The rest of the article is organized as follows. We start by explaining the concept of trust in fairness for spectator sports and briefly review the related literature. Then, we describe the data and our study samples. This is followed by analyses of the panel survey and the scenario. The final section concludes.

Trust in fairness and doping

Our study follows the assumption that consumer preferences for elite sports are based on two inherent features of sporting competitions: top-level athletic performances and fairness of the competition, also described as compliance with the rules or integrity of sports (Buechel et al., 2016; Mills, 2017). Both features are consumed simultaneously and expected to be essential for producing suspense and outcome uncertainty in sporting competitions.[2] However, athletes' unethical and illegal conduct such as doping is usually hidden and mostly revealed *after* a competition – quite often with a considerable time lag. Sport consumers therefore need to trust that athletes will compete fairly. In this context,

[2] For a discussion and some recent evidence on the (ir)relevance of suspense and outcome uncertainty in spectator sports; see Coates et al. (2014), Pawlowski (2013), and Pawlowski et al. (2018).

top athletes may theoretically act as (dis)trust ambassadors signaling the (un)fairness of the competition (McEvily et al., 2003).

In general, trust is a complex construct that describes the trustor's willingness to show vulnerability to others due to the expectation that those other's intentions and conduct are favorable in turn (Mayer et al., 1995). It further distinguishes between the trusting intention and trusting beliefs in the other's ability, benevolence, and integrity in a given situation (McKnight et al., 1998). The latter represents trust in the moral and ethical conduct of the trustee and is considered particularly important in the context of elite sports (Gardiner et al., 2017). As such, it is a popular claim that doping scandals reduce the demand for sports driven by a loss of trust in fairness (Dimant & Deutscher, 2019; Frenger et al., 2013). We illustrate this relationship in Figure 1. However, neither the effect of doping scandals on demand nor the supposed underlying mechanism via trust are well established in the literature.

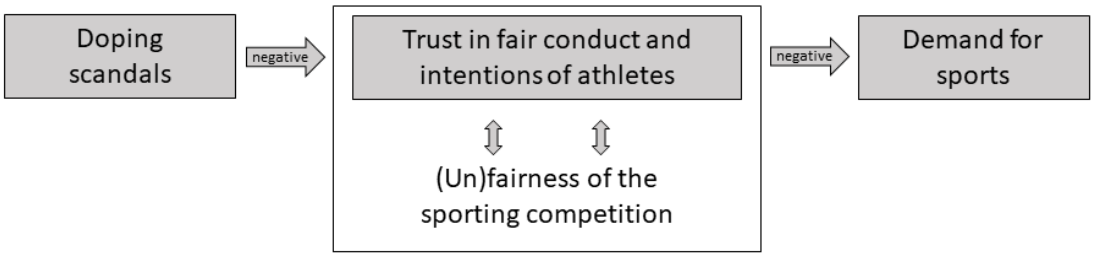

Figure 1. Conceptualizing the relationship between doping, trust in fairness, and the demand for sports.

For instance, Cisyk and Courty (2017) and Cisyk (2020) found that revealed PED suspensions in major league baseball (MLB) have a negative short-term effect on home-game attendance and the TV audience for the affected team, while spillover effects across MLB teams are only found for game attendance. In contrast, Brave and Roberts (2019) did not find any impact of PED suspensions on gate revenues per season in MLB. Likewise, studies on cycling found negative long-term effects of doping scandals on the TV audience in Flanders, Belgium (Van Reeth, 2013) but no long-term effects in Spain (Rodríguez et al., 2015). Additionally, Abeza et al. (2020) indicate that doping is regarded as a serious offense, altering consumer preferences for watching the Australian Football League, whereas Hallmann et al. (2017) did not find any correlation between the perceived number of doped athletes and the willingness to support elite sports.

There are some plausible reasons why previous empirical studies on the doping–demand relation have been inconclusive. For instance, consumers may also derive some utility from witnessing scandals and listening to discussions on doping news. This suggests that they keep following their sport in order to stay updated about how doping cases unfold (Van Reeth, 2013). As such, doping might not necessarily reduce demand. Moreover, according to the economics of superstars, well-performing athletes drive demand by their superior talent and quality of play (MacDonald, 1988; Rosen, 1981). Therefore, lower demand might also be a response to reduced performance quality if a top athlete is banned because of doping. This is in line with arguments provided by Buraimo et al. (2016) explaining why game attendance decreased after the Calciopoli scandal in Italian soccer. They also provide some empirical evidence that the decrease in demand following corruption is not driven by the consumers' moral disapproval. Likewise, Manoli et al. (2020) also question the role of trust in fairness as an integral part of the overall sports system.

In summary, the inconsistent findings suggest that the doping–demand relation is not as straightforward as often claimed. In addition, previous literature lacks a direct test of the potential main driver of reduced sports demand after scandals of unethical or illegal conduct, i.e. trust in fairness. Testing this potential mechanism with secondary data appears to be difficult (e.g. Cisyk & Courty, 2017), which is why we use and combine primary data collected from a panel survey and a scenario approach as described in the following.

Data and sample

We use data from a large two-wave online survey that was conducted by a market research company in Germany before and after three major sporting events in summer 2018: the FIFA World Cup, the Tour de France, and the European Athletics Championships (EAC). The first wave took place from the 28th of May to the 11th of June 2018. The second wave started on the 13th of August and ended on the 24th of August 2018. The sample is representative for the adult German population. The survey contains a rich set of questions on T&F, cycling, and soccer, as well as general and sport-related measures of

trust.[3] In our paper, we make use of the panel sample, which consists of the same individuals in both waves, to examine the role of trust in fairness for spectator sports demand *in general* (see Panel survey analysis). In order to further investigate whether unethical and illegal conduct cause a loss of trust followed by reduced sports demand *in particular*, we intended to exploit a 'doping shock' in a field experiment at the EAC. However, since we did not identify any serious doping scandals during our observation window that may have led to a loss of trust, we exploit a sporting event scenario where doping issues are present. The scenario was presented in the second survey wave to a randomly drawn subsample of the panel sample (see Scenario analysis).

After imposing quality corrections and excluding observations with missing values, our panel sample consists of 1,756 individuals; the scenario subsample includes 271 individuals. As claimed before, both samples have a similar distribution with regard to age, gender, and region compared to the adult German population (see Table A1, Online Appendix A).

Panel survey analysis

In this first part, we examine the role of trust for the TV demand of the EAC 2018. We consider this T&F event an ideal setting because the sport has been affected by several doping scandals in the past. The EAC was part of the inaugural edition of the European Championships and staged in Berlin from the 6th to the 12th of August 2018. The event was broadcasted via the Eurovision Network across Europe and watched by a large audience (e.g. 5.18 million viewers on German TV at peak times, 18% market share).[4]

Measures

Our outcome variable *EAC 2018 TV consumption* was surveyed in wave two, i.e. after the event, and measures the number of competition days watched by the study participants. In contrast to measuring attendance of major sporting events, which involves high opportunity costs (e.g. travel costs), using TV consumption to approximate spectator demand for such events has the advantage to obtain a heterogenous sample of viewers

[3] The other parts of the survey were used for a study on following athletes and teams on social media (Utz et al., 2021) and a crisis communication experiment, both thematically unrelated to our paper and using a different sample.
[4] TV rating data are taken from the AGF Videoforschung GmbH (2018).

and non-viewers. In addition, each of the six EAC competition days was broadcast live on one of the major public free-TV channels (Das Erste and ZDF) and on the sports channel Eurosport. As such, these broadcasts were easily accessible for the German population.

Our trust variables were measured in wave one, i.e. before the event. In our setting, we considered the athletes as key trust agents who signal the (un)fairness of the sporting competition to consumers. The *trusting intention* in athletes as well as the trusting belief in their integrity, i.e. the *honesty* and *fairness* of athletes (Mayer et al., 1995), are combined into the index variable *trust in T&F athletes* (Cronbach's $\alpha = 0.95$; measures adapted from Dieckmann et al., 2016) in order to avoid issues of multicollinearity in the subsequent analysis.[5] Following McKnight et al. (1998), we also control for differences in trust formation due to institution-based trust and disposition to trust. As before, trust in the institutions responsible for organizing fair competitions at the national and international level, i.e. *trust in the German Athletics Association* (GAA) and trust in the *International Association of Athletics Federations* (IAAF), are combined into an index variable, measuring *institution-based trust* (Cronbach's $\alpha = 0.88$). The variable *generalized trust* (Cronbach's $\alpha = 0.78$) controls for the general disposition to trust (Mayer et al., 1995; items were taken from Beierlein et al., 2012).

In order to account for any confounding influences in our analysis, we control for sport-related features as well as socio-demographics that were previously found to influence demand. Like all trust measures, these control variables were measured in wave one. As such, we control for the participants' *general interest* to watch T&F events to consider the intrinsic preference for this sport (Borland & MacDonald, 2003)[6] as well as *patriotism*, which may drive demand of sporting events with a national character and compatriot athletes (Nüesch & Franck, 2009). Next to these context-specific variables, we control for *gender*, *age*, being *employed*, the level of *education* and *income*, as well as the *relationship status* to proxy differences in individual preferences and account for

[5] To further assess multicollinearity issues in our data, we consulted correlation coefficients, variance-inflation factors, and looked at suspicious changes in the model outputs.
[6] T&F events regularly involve multiple disciplines that also overlap in their temporal order. As such, measuring *directly* common sport-specific features (e.g., quality of the competition) is not feasible.

budget and time constraints (e.g. Nalbantis & Pawlowski, 2019; Pawlowski et al., 2018; see Table A2, Online Appendix A for variable descriptions).

Table 1 presents the descriptive statistics of all variables. Participants watched on average nearly two competition days on TV (45% of the sample did not watch the EAC at all). They also show moderate levels of trust and *interest in T&F*. 57% of the sample are *male* and the average *age* is 53 years. Almost 50% of the participants have a medium income and most of them are in a relationship.

Table 1. Descriptive statistics of the variables.

Variables	Mean	SD	Min	Max
EAC 2018 TV consumption	1.94	2.20	0	6
Trust in T&F athletes	3.40	0.92	1	5
Institution-based trust	3.17	0.98	1	5
Generalized trust	3.16	0.81	1	5
Interest in T&F	2.73	1.31	1	5
Patriotism	3.71	1.04	1	5
Male	0.57	0.50	0	1
Age	53.10	12.09	18	78
Employed	0.59	0.49	0	1
Education				
Other	0.64	0.48	0	1
Upper secondary school	0.17	0.38	0	1
University degree	0.18	0.39	0	1
Income				
<2,000€	0.36	0.48	0	1
2,001–4,000€	0.49	0.50	0	1
>4,001€	0.15	0.36	0	1
Relationship				
married	0.55	0.50	0	1
Single	0.16	0.37	0	1
Living with partner	0.15	0.36	0	1
Divorced/widowed	0.14	0.34	0	1
Number of obs.		1,756		

Empirical strategy

We regress *EAC 2018 TV consumption* on *trust in T&F athletes* as well as the other trust measures and confounding variables as described before. By exploiting the panel structure of our data, i.e. by taking all predictor variables from the first wave and the outcome variable from the second wave, we are able to purge any issues of common method variance and reverse causality (Antonakis et al., 2010). Given the nature of our dependent variable, we employ count data models next to ordinary least squares (OLS) regressions. We opt for the ZIP model after comparing the Poisson model, the negative binomial model (NBM), the ZIP model, and the zero-inflated NBM based on the Akaike

Information Criterion (AIC), the Bayesian Information Criterion (BIC), and likelihood-ratio tests.

The ZIP accounts for the large excess of zeros in our outcome variable (45% of the sample) and allows zero consumption to be generated by two different processes, meaning the existence of two subgroups (Lambert, 1992): The zero-consumption group does not even contemplate watching the EAC, which means that these individuals always have zero-consumption with a probability of one. This could be the case if, for example, people may not have a TV at home or generally do not watch any sports. The potential-consumption group has a non-zero probability of a positive outcome, which means that these individuals may (not) consume the EAC depending on the conditions. Therefore, ZIP models consist of two parts. A logit model estimates the probability if an individual belongs to the zero-consumption group and a Poisson model estimates the probability of counts (or frequency) of EAC consumption days (including possible zeros). In order to facilitate easy interpretation of the effects, we present the incidence rate ratios (IRR) and the odds ratios (OR) from the estimated coefficients of the Poisson and logit model, respectively (Muñiz et al., 2014). The IRR is the relative change in the incidence rate for a one-unit change in any given variable, holding the other predictor variables constant. The OR is the relative change in the odds of zero consumption for a one-unit increase in the predictor variable. Values of the IRR and OR greater (smaller) than one indicate that the predictor variable has a positive (negative) effect.

Results

The results are presented in Table 2. In Column 1, we regress our demand outcome on *trust in T&F athletes*. As expected, consumer trust positively correlates with the level of TV consumption. This relationship remains significant after adding *institution-based trust* and *generalized trust* to control for differences in trust formation (Column 2, Table 2). However, the relation of interest becomes insignificant after controlling for sport-related and socio-demographic factors (Column 3, Table 2). This finding is supported by our ZIP estimations (Column 5, Table 2) since our trust measures neither significantly affect the likelihood of an individual to consume the EAC nor the frequency of TV consumption.

Table 2. Results of OLS and ZIP regressions on the EAC 2018 TV consumption.

Variables	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) ZIP	(6) ZIP
					potential consumption	
Trust in T&F athletes	**0.431 (0.000)**	**0.289 (0.000)**	0.040 (0.511)	0.053 (0.683)	0.985 (0.535)	1.012 (0.646)
Institution-based trust		0.110 (0.111)	-0.081 (0.146)	-0.012 (0.922)	0.985 (0.483)	1.019 (0.440)
Generalized trust		**0.200 (0.005)**	0.026 (0.642)	0.105 (0.389)	1.010 (0.667)	1.017 (0.493)
Interest in T&F			**0.984 (0.000)**		**1.208 (0.000)**	
Patriotism			0.069 (0.123)	0.026 (0.791)	**1.042 (0.043)**	**1.071 (0.002)**
Male			**0.206 (0.013)**	**0.349 (0.055)**	1.047 (0.179)	1.057 (0.138)
Age			**0.020 (0.000)**	**0.041 (0.000)**	**1.008 (0.000)**	**1.010 (0.000)**
Employed			**-0.285 (0.002)**	**-0.364 (0.067)**	**0.910 (0.015)**	**0.902 (0.013)**
Education (ref.: other)						
Upper secondary school			0.028 (0.812)	0.039 (0.876)	1.068 (0.141)	**1.087 (0.088)**
University degree			-0.021 (0.848)	0.074 (0.751)	0.992 (0.863)	1.012 (0.810)
Income (ref.: <2,000€)						
2,001–4,000€			0.097 (0.342)	-0.122 (0.594)	0.999 (0.977)	1.000 (0.997)
>4,001€			0.186 (0.197)	0.014 (0.965)	0.985 (0.792)	0.976 (0.685)
Relationship (ref: married)						
Single			0.106 (0.414)	0.399 (0.179)	1.054 (0.359)	1.094 (0.144)
Living with partner			0.057 (0.657)	0.258 (0.328)	1.073 (0.177)	1.084 (0.147)
Divorced/widowed			**-0.240 (0.079)**	-0.151 (0.618)	1.027 (0.617)	1.023 (0.696)
					zero consumption	
Trust in T&F athletes					0.880 (0.185)	0.901 (0.284)
Institution-based trust					1.114 (0.238)	1.139 (0.163)
Generalized trust					0.983 (0.857)	0.991 (0.923)
Interest in T&F					**0.291 (0.000)**	**0.264 (0.000)**
Patriotism					0.957 (0.563)	0.954 (0.551)
Male					**0.752 (0.039)**	**0.760 (0.050)**
Age					**0.987 (0.056)**	**0.988 (0.079)**
Employed					1.173 (0.297)	1.176 (0.297)
Education (ref: other)						
Upper secondary school					1.220 (0.288)	1.222 (0.295)
University degree					0.949 (0.770)	0.952 (0.786)
Income (ref.: <2,000€)						
2,001–4,000€					0.806 (0.202)	0.808 (0.214)
>4,001€					**0.602 (0.025)**	**0.613 (0.032)**
Relationship (ref: married)						
Single					1.037 (0.869)	1.069 (0.767)
Living with partner					1.132 (0.550)	1.162 (0.475)
Divorced/widowed					**1.546 (0.050)**	**1.536 (0.059)**
R^2/Wald χ^2	0.033	0.039	0.399	0.082	220.82 (0.000)	88.58 (0.000)
Number of obs.	1,756	1,756	1,756	539	1,756	1,756

Notes: Presented are the unstandardized coefficients for the OLS estimations. In the ZIP models, we present the incidence rate ratios for the Poisson estimations (potential consumption) and the odds ratios for the logit estimations (zero consumption). Robust standard errors are used in all models. *P*-values are presented in parentheses. Bold numbers indicate significance at conventional levels ($p < 10\%$).

Regarding the full model specification, a concern is that using *interest in T&F* as a control may lead to biased estimates because this variable might be affected by our variable of interest. By estimating an OLS model with a subsample of interested individuals (*interest in T&F* > 3) and excluding *interest in T&F* (Column 4, Table 2), we are able to test the relation of interest only for those who would generally like to watch T&F events. In addition, we make use of the flexibility of the ZIP and estimate a model that uses *interest in T&F* to only predict whether the individuals' TV consumption is always zero (see zero-consumption estimation in Column 6, Table 2). Again, both models show no statistically significant effects of trust.

Scenario analysis

Although we do not find any significant effect of trust in fairness in our panel survey analysis, the question remains if a substantial loss of consumer trust may impact demand after scandals of unethical and illegal conduct. We therefore make use of a real doping case and a sporting event scenario to further investigate whether such scandals cause a loss of trust followed by reduced demand.

Scenario description

Our scenario presents a real contest of an upcoming international T&F event, i.e. the men's final of the 100-meter dash at the Internationale Stadionfest (ISTAF) meeting in Berlin in 2018, in which a top-performing athlete with PED abuse in the past is added to the starting list. The 100-meter dash is considered one of the most prestigious and popular events in T&F.

In the scenario text, we first describe the event (type, date, and location) and then list the participating athletes. We manipulated the starting list of the competition and included all athletes of the recently held 100-meter final of the EAC 2018. These athletes have never been associated with any doping issues in their careers. In addition, we *explicitly* added the incumbent World Champion, Justin Gatlin, to the field of competitors. The US sprinter won the World Championships for the second time in 2017 after his first victory in 2005. He also won gold, silver, and bronze medals in the 100-meter dash at various Olympic Games. Besides his outstanding performances and glory, he was convicted of doping twice in his career. This took place first in 2001, when he was suspended for two years due to the use of amphetamines. This ban was reduced to one year by the IAAF after Gatlin appealed the decision. He was convicted again for the use of exogenous testosterone after a positive doping test in 2006 and was initially banned for eight years (reduced to four years in 2007; American Arbitration Association, 2008). His consecutive PED abuse has been discussed and much criticized in the media and made him an unwelcome person in professional T&F events.[7]

[7] For instance, see The Telegraph article for a media discussion: https://www.telegraph.co.uk/athletics/2018/02/10/cannot-stop-justin-gatlin-running-london-admits-world-cup-chief/. Retrieved March 15, 2021.

The scenario was presented to a randomly drawn subsample of our panel sample. To identify the relation of interest, we compare individuals who are aware of the Justin Gatlin doping case with those who are not. The intuition is that Gatlin acts as a (dis)trust ambassador for all athletes in this setting and we expect that well-informed people have experienced a substantial loss of trust in fairness. As such, we define awareness of the Gatlin case as the treatment condition in our estimation approach.

Outcome measures

As in our panel survey analysis, *trust in competitors* measures the trusting intentions in all participating athletes while *honesty* and *fairness* of competitors measure the trusting belief of integrity. Additionally, the variable *trust in no doping* specifically measures the level of trust that the athletes behave fairly, i.e. do not violate the anti-doping rules.

We use an *intention-to-consume* (ITC) as well as a *willingness-to-pay* (WTP) measure as demand proxies (e.g. Nalbantis et al., 2017; Pawlowski et al., 2018). Our ITC measure is a dummy that indicates whether a respondent intends to watch the 100-meter dash live on free TV. We further correct for any hypothetical bias by using *ex post* certainty calibration (Loomis, 2011). If respondents stated that their certainty of actual consumption is greater than 70% (see Morrison & Brown, 2009), our bias-corrected (bc) measure receives the value of one, and zero otherwise.

In general, evaluating the WTP of watching the 100-meter dash live on TV is a difficult task for respondents because major T&F events are commonly broadcasted on public free TV in Germany. However, since we are only interested in comparing the WTP between the *treated* (aware) and *untreated* (unaware) groups, we argue that a direct approach with an open-ended elicitation format is valid (van Doorn & Verhoef, 2011). To set boundaries and reduce value uncertainty, we use the fee of 210€ that each German household has to pay per annum to the public broadcasting services. More precisely, we ask respondents how much of this public broadcasting fee they would be willing to spend to ensure that this event is broadcasted live on public TV (see Table A2, Online Appendix A for variable descriptions).

Table 3 presents the conditional mean values of all trust and demand measures. As can be seen, the *treated* (T) and *untreated* (U) groups differ significantly with regard to the level of trust as well as both ITC measures (see Columns 1, 2, and 4, Table 3). On average, mean values for the trust measures are significantly lower while means for both ITC measures are significantly higher for the *treated* group.

We also compare the *untreated* group (U) – which is unaware of the doping case but knows that Gatlin is a superstar – with 156 individuals from another randomly drawn subsample who were asked the same questions about the event, though only with athletes of the recently held 100-meter final of the EAC 2018 as competitors (WG; see Column 3, Table 3). It can be seen that both groups (U and WG) do not differ in the outcome variables (Column 5, Table 3). We therefore argue that any differences between T and U are not driven by including a top athlete and superstar rather than any mediocre athlete in our sporting event scenario. Moreover, we argue that using the (un)awareness of the Gatlin case as the treatment condition to examine the impact of doping on consumers' trust and demand seems to be a reasonable approach.

Table 3. Mean values of the outcome variables and t-test significance.

Outcome variable	(1) Treated (T)	(2) Untreated (U)	(3) Without Gatlin (WG)	(4) *t*-test T-U	(5) *t*-test U-WG
Trust in competitors	2.70	3.33	3.27	**-3.90 (0.000)**	0.52 (0.605)
Honesty of competitors	2.66	3.29	3.22	**-3.81 (0.000)**	0.69 (0.490)
Fairness of competitors	3.02	3.51	3.46	**-2.96 (0.003)**	0.50 (0.617)
Trust in no doping	2.56	3.16	3.23	**-3.28 (0.001)**	-0.54 (0.589)
Intention-to-consume	0.70	0.49	0.54	**2.67 (0.008)**	-0.86 (0.388)
Intention-to-consume (bc)	0.48	0.30	0.31	**2.48 (0.014)**	-0.32 (0.749)
Willingness-to-pay	26.91	37.21	36.98	-1.35 (0.179)	0.04 (0.966)
Number of obs.	50	221	156		

Notes: *P*-values of the *t*-tests in parentheses. Bold numbers indicate significance at conventional levels ($p < 10\%$).

Empirical strategy

Since being aware (or not) of the Gatlin case is not a random process and depends on other factors such as a general interest in T&F, we cannot derive any causal interpretation from just comparing the mean values between the *treated* and the *untreated* groups. In order to address such endogeneity and selection issues in our scenario setting, we implement a selection-on-observable approach. More precisely, to ensure that the conditional independence assumption (CIA) holds (Imbens, 2004), we control for all

variables jointly affecting awareness of the Gatlin case as well as our trust and demand outcomes.

As such, we use the predictors of our survey analysis as confounding variables, except for *patriotism* because the scenario does not include any compatriot athlete. Instead, we control for *expertise in T&F*,[8] which, together with *interest in T&F*, is considered a precondition for awareness of doping issues and for developing some form of trust in the involved agents (Lewicki & Bunker, 1995). *Trust in T&F athletes, institution-based trust*, and *generalized trust* are used as trust antecedents to control for differences in trust formation and trusting beliefs. Next to these context-specific variables, we control for socio-demographics.

In order to avoid introducing any endogeneity issues, we exploit the panel structure of our data by taking the confounding variables from the first wave of data collection (except for *expertise in T&F*), while the treatment and outcome variables are measured in the second wave. Finally, the order of questions in the second wave is used to avoid any response bias: after the scenario description, respondents first provided demand-related information, followed by questions on the trust outcomes. Then, more fact-based questions on awareness of the Gatlin case and the expertise were asked. Asking about awareness in the latter part of the questionnaire ensured that respondents were not primed to the issue of doping.

To analyze the effects of the doping case, we use radius matching with bias adjustment based on the propensity score (Lechner et al., 2011). In a first step, a probit model is used with all confounding variables as predictors to estimate the propensity score. In our study, the propensity score defines the probability of being aware of the Gatlin case. In a second step, *(un)treated* observations are matched with all comparison observations in a predefined radius around the propensity score. Observations that cannot be matched due to a lack of common support in the propensity score are automatically excluded from the analysis. We report the average treatment effect (ATE), the average treatment effect on the treated (ATET), and the average treatment effect on the non-treated (ATENT). The ATE is the expected effect for a randomly selected individual from the population. The

[8] To measure *expertise in T&F*, we asked for the time of the current world record in the 100-meter dash (set by Usain Bolt in 9.58 s). The variable is a distance measure stating the absolute difference between estimated and actual world record.

ATET and ATENT measure the expected effects in the *treated* and *untreated* subpopulations, respectively (Imbens & Wooldridge, 2009). Boot-strapping (999 times) is used for statistical inference.

Overall, we use two different specifications in our analysis. The first model includes all sports-related and socio-demographic characteristics as confounding variables and the trust and demand measures as outcome variables. The second model additionally controls for all trust antecedents in the matching process.

Results

Table 4. Mean values and marginal effects of the selection models.

	Mean values in subsamples		Marginal effects	
	(1)	(2)	(3)	(4)
Variable	Treated	Untreated	Model 1	Model 2
Interest in T&F	3.48	2.56	**0.061 (0.013)**	**0.057 (0.024)**
Expertise in T&F (w2)	0.45	2.65	-0.042 (0.624)	-0.116 (0.260)
Male	0.72	0.49	0.097 (0.131)	0.051 (0.381)
Age	53.54	51.84	-0.002 (0.354)	-0.003 (0.284)
Employed	0.58	0.60	-0.039 (0.487)	-0.077 (0.287)
Education (ref.: other)	0.60	0.68		
Upper secondary school	0.16	0.19	-0.015 (0.819)	0.001 (0.986)
University degree	0.24	0.13	0.113 (0.160)	0.114 (0.186)
Income (ref.: <2,000€)	0.26	0.36		
2,001–4,000€	0.62	0.48	0.097 (0.145)	**0.147 (0.040)**
>4,001€	0.12	0.16	0.037 (0.656)	0.109 (0.254)
Relationship (ref.: married)	0.42	0.51		
Single	0.22	0.20	0.047 (0.498)	0.074 (0.386)
Living with partner	0.18	0.15	0.075 (0.340)	0.063 (0.470)
Divorced/widowed	0.18	0.14	0.149 (0.131)	**0.214 (0.063)**
Trust in T&F athletes [a)]	3.55	3.50		0.037 (0.401)
Institution-based trust [b)]	3.11	3.28		-0.044 (0.321)
Generalized trust	3.36	3.15		0.040 (0.343)
Number of obs./Efron's R^2	50	221	271 / 0.193	229 / 0.244

Notes: "w2" = variable is measured in the second survey wave (all other variables are measured in the first wave). Deviations in the number of observations occur due to "don't know" answers (treated/untreated): [a)] 47/201, [b)] 47/197. The dependent variable in the probit models is a dummy measuring one if respondents are aware of the Gatlin case. Presented are the average marginal effects. *P*-values (in parentheses) obtained from 999 bootstrap replications. Bold numbers indicate significance at conventional levels ($p < 10\%$).

Table 4 provides an overview of the conditional mean values and the average marginal effects of all variables used to model the selection process. As expected, the *treated* group has on average more *interest in T&F* and a better *expertise in T&F* (Columns 1 and 2, Table 4).[9] However, any differences in the conditional mean values in both subsamples

[9] Note that a lower value in *expertise in T&F* corresponds to a higher level of expertise.

are apparently washed out once we control for *interest in T&F* (Columns 3 and 4, Table 4). This is confirmed by the balancing statistics of both models. Overall, the mean biases are considerably reduced for the matched samples (see Table B1 and B2, Online Appendix B). Therefore, we argue that conditioning on the confounding variables was successful.

Table 5. Effects of the Gatlin case.

	Model 1 [a)]			Model 2 [b)]		
Outcome variables	ATE	ATET	ATENT	ATE	ATET	ATENT
Trust in competitors	**-0.695**	**-0.892**	**-0.647**	**-0.805**	**-0.659**	**-0.859**
	(0.001)	**(0.007)**	**(0.002)**	**(0.003)**	**(0.067)**	**(0.010)**
Honesty of competitors	**-0.450**	**-0.785**	**-0.367**	**-0.838**	**-0.678**	**-0.898**
	(0.038)	**(0.009)**	**(0.086)**	**(0.002)**	**(0.071)**	**(0.008)**
Fairness of competitors	-0.114	**-0.647**	0.019	**-0.837**	**-0.644**	**-0.909**
	(0.736)	**(0.010)**	(0.973)	**(0.003)**	**(0.026)**	**(0.015)**
Trust in no doping	**-0.446**	**-0.990**	-0.310	**-0.739**	**-0.744**	**-0.737**
	(0.050)	**(0.006)**	(0.245)	**(0.015)**	**(0.067)**	**(0.031)**
Intention-to-consume	-0.085	-0.085	-0.085	-0.111	-0.067	-0.127
	(0.370)	(0.436)	(0.318)	(0.363)	(0.595)	(0.423)
Intention-to-consume (bc)	-0.038	-0.098	-0.023	-0.013	0.014	-0.024
	(0.656)	(0.421)	(0.798)	(0.910)	(0.944)	(0.846)
Willingness-to-pay	32.250	-1.199	40.567	**-15.271**	-13.730	**-15.840**
	(0.174)	(0.900)	(0.166)	**(0.061)**	(0.174)	**(0.057)**

Notes: Presented are the treatment effects (ATE, ATET, & ATENT) of being aware of the Gatlin case. [a)] number of observations = 271, common support = 231 (85%), number of treated = 46; [b)] number of observations = 229, common support = 152 (66%), number of treated = 41. *P*-values (in parentheses) obtained from 999 bootstrap replications. Bold numbers indicate significance at conventional levels ($p < 10\%$).

All treatment effects (ATE, ATET, and ATENT) are presented in Table 5. Results of our first model show a significant negative effect of being aware of the Gatlin case on *trust in competitors*. The effects on the integrity measures are also negative, though the ATE is only significant for *honesty of competitors*. Specific trust that the athletes do not violate the anti-doping rules significantly decreases as well. Importantly, however, we do not find any impact on demand. Compared with our first model, the second model, which additionally conditions on the trust antecedents, indicates much weaker common support. However, since off-support is mainly driven by unmatched observations in the *untreated* group, we report these results as well. Overall, the average effects on all trust outcomes are negative and significant (see model 2, Table 5). Moreover, while the effects on both intention-to-consume outcomes remain insignificant, the effect on *willingness-to-pay* becomes negative and marginally significant. However, the mean values of the *treated* and *untreated* groups indicate that a certain proportion of uninformed individuals exists

without having any propensity to watch a T&F event on TV. Therefore, we rerun our models with a subsample that excludes respondents with the lowest level of *interest in T&F*. While our main findings remain, the marginally significant effects on *willingness-to-pay* diminish (see Table C2, Online Appendix C).

Conclusion

We make use of a two-wave panel survey and a sporting event scenario to examine the impact of trust in fairness on spectator sports demand *in general* and its role after a doping scandal *in particular*. In the first part of our analysis, we regress the TV consumption of a major T&F event on a trust in fairness measure. In the second part, we exploit a scenario with a real doping case presented to a randomly drawn subsample in the second survey wave.

Our regression results suggest that trust in fairness has no significant impact on TV consumption of an international T&F event. Likewise, while results of our radius matching estimations show that awareness of a major doping case reduces the average level of trust in athletes' fair conduct and integrity, we do not find a (negative) effect on the ITC and the WTP in terms of watching the affected event. As such, our findings suggest that anti-doping rule violations by an athlete may unfold a long-lasting effect on consumers' trust in fairness, which confirms theoretical assumptions in related work (Dimant & Deutscher, 2019; Frenger et al., 2013). However, our findings generally question the role of trust in fairness for spectator sports demand. Even a substantial loss of trust after a major doping case does not necessarily reduce demand, which suggests that fairness is not, or at least less relevant for consumers to keep following sports. Although trust was found to drive demand in different economic settings (e.g. Ding et al., 2013; Guiso et al., 2008), this conclusion is consistent with findings from other sport-related studies (Buraimo et al., 2016; Manoli et al., 2020).

Our study adds to the ongoing discussion about the doping–demand relation and its potential mechanisms. In this regard, our results may help to explain why studies could not observe any negative long-term effects after news on individual athletes' PED abuse (Cisyk, 2020; Cisyk & Courty, 2017) and doping scandals (Rodríguez et al., 2015). Sports fans may quickly return to their previous consumption habits, although they are likely to

have experienced a (long-term) loss of trust in the fair conduct and integrity of athletes. Our study also supports the notion of Van Reeth (2013) that public discussions on doping could be part of professional sports entertainment. Moreover, the different findings across doping–demand studies may also reflect the level of interest for doping and fairness across the analyzed sports and countries where the studies were conducted.

Our findings also have practical implications. On the one hand, managers can be less concerned that individual doping cases and a loss of trust in fairness have serious long-term consequences for sports demand. On the other hand, however, sports policy makers should be concerned about PED abuse if they use elite sports as a vehicle to promote social values such as moral principles and fair play. Previous findings suggest that trust in the relevant actors is important for the public evaluation of elite sports and related sport policy programs (Funahashi et al., 2015; Hallmann et al., 2020). We add to this discussion by observing a loss of trust in fairness after a major doping case. This finding might serve as an argument to justify anti-doping policies with the aim of preserving elite sports as a platform for ethical conduct and fairness as suggested in related work on the societal impact of elite sports (e.g. De Rycke & De Bosscher, 2019, 2020). It may also be important for other stakeholders such as sponsors who intend to benefit from the positive social values of elite sports.

However, our study has also certain limitations. First, our scenario analysis is based on stated preferences instead of actual consumer choices. Second, different doping scandals in sports may be evaluated differently by individuals and the public. Therefore, future research should test the external validity of our survey and scenario findings by examining other sports and doping scandals. This may include, for instance, less well-known or compatriot athletes who are convicted of PED abuse, or doping scandals that occur at the team or country level.

References

Abeza, G., O'Reilly, N., Prior, D., Huybers, T., & Mazanov, J. (2020). The impact of scandal on sport consumption: Do different scandal types have different levels of influence on different consumer segments? *European Sport Management Quarterly*, *20*(2), 130–150. https://doi.org/10.1080/16184742.2019.1586974

AGF Videoforschung GmbH. (2018, August 12). Marktdaten [Market data]. Retrieved August 14, 2018, from https://www.agf.de/daten/tv-daten

American Arbitration Association (AAA/North American Court of Arbitration for Sport). (2008). *AAA No. 30 190 00170 07*. Retrieved December 10, 2019, from https://www.usada.org/wp-content/uploads/AAA-CAS-Decision-Justin-Gatlin-January-2008.pdf

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010

Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2012). Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen (KUSIV3) [Scale to measure interpersonal trust]. *GESIS-Working Papers*.

Borland, J., & MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, *19*(4), 478–502. https://doi.org/10.1093/oxrep/19.4.478

Brave, S. A., & Roberts, K. A. (2019). The competitive effects of performance-enhancing drugs: MLB in the posttesting era. *Journal of Sports Economics*, *20*(6), 747–781. https://doi.org/10.1177/1527002518817584

Buechel, B., Emrich, E., & Pohlkamp, S. (2016). Nobody's innocent: The role of customers in the doping dilemma. *Journal of Sports Economics*, *17*(8), 767–789. https://doi.org/10.1177/1527002514551475

Buraimo, B., Migali, G., & Simmons, R. (2016). An analysis of consumer response to corruption: Italy's Calciopoli scandal. *Oxford Bulletin of Economics and Statistics*, *78*(1), 22–41. https://doi. org/10.1111/obes.12094

Chanley, V. A., Rudolph, T. J., & Rahn, W. M. (2000). The origins and consequences of public trust in government: A time series analysis. *Public Opinion Quarterly*, *64*(3), 239–256. https://doi.org/10.1086/317987

Cisyk, J. (2020). Impacts of performance-enhancing drug suspensions on the demand for Major League Baseball. *Journal of Sports Economics*, *21*(4), 391–419. https://doi.org/10.1177/1527002520906529

Cisyk, J., & Courty, P. (2017). Do fans care about compliance to doping regulations in sports? The impact of PED suspension in baseball. *Journal of Sports Economics*, *18*(4), 323–350. https://doi.org/10.1177/1527002515587441

Coates, D., Humphreys, B. R., & Zhou, L. (2014). Reference-dependent preferences, loss aversion, and live game attendance. *Economic Inquiry*, *52*(3), 959–973. https://doi.org/10.1111/ecin.12061

De Rycke, J., & De Bosscher, V. (2019). Mapping the potential societal impacts triggered by elite sport: A conceptual framework. *International Journal of Sport Policy and Politics, 11*(3), 485–502. https://doi.org/10.1080/19406940.2019.1581649

De Rycke, J., & De Bosscher, V. (2020). The cure or the cause? Public opinions of elite sports' *societal benefits and harms*. *Sport in Society*, 1–20. https://doi.org/10.1080/17430437.2020.1724962

Dieckmann, A., Grimm, V., Unfried, M., Utikal, V., & Valmasoni, L. (2016). On trust in honesty and volunteering among Europeans: Cross-country evidence on perceptions and behavior. *European Economic Review*, *90*, 225–253. https://doi.org/10.1016/j.euroecorev.2016.01.011

Dimant, E., & Deutscher, C. (2019). The economics of doping in sports—a special case of corruption. In P. Downward, B. Frick, B. R. Humphreys, T. Pawlowski, J. E. Ruseski, & B. P. Soebbing (Eds.), *The SAGE handbook of sports economics* (pp. 544–552). Sage.

Ding, Y., Veeman, M. M., & Adamowicz, W. L. (2013). The influence of trust on consumer behavior: An application to recurring food risks in Canada. *Journal of Economic Behavior & Organization*, *92*, 214–223. https://doi.org/10.1016/j.jebo.2013.06.009

European Commission. (2007). *White Paper on Sport* (Report No. COM (2007) 391 final). Federal Ministry of Justice and Consumer Protection (Bundesministerium der Justiz und für Verbraucherschutz). (2015). *Act Against Doping in Sport*. Retrieved December 10, 2019, from http://www.gesetze-im-internet.de/englisch_antidopg/

Frenger, M., Emrich, E., & Pitsch, W. (2013). How to produce the belief in clean sports which sells. *Performance Enhancement & Health*, *2*(4), 210–215. https://doi.org/10.1016/j.peh.2014.09.001

Funahashi, H., De Bosscher, V., & Mano, Y. (2015). Understanding public acceptance of elite sport policy in Japan: A structural equation modelling approach. *European Sport Management Quarterly*, *15*(4), 478–504. https://doi.org/10.1080/16184742.2015.1056200

Gardiner, S., Parry, J., & Robinson, S. (2017). Integrity and the corruption debate in sport: Where is the integrity? *European Sport Management Quarterly*, *17*(1), 6–23. https://doi.org/10.1080/16184742.2016.1259246

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance*, *63* (6), 2557–2600. https://doi.org/10.1111/j.1540-6261.2008.01408.x

Hallmann, K., Breuer, C., Ilgner, M., & Giel, T. (2017). Public perceptions on the dark side of elite sports and its influence on the willingness to support elite sports. In

B. G. Pitts & J. J. Zhang (Eds.), *Global sport management: Contemporary issues and inquiries* (pp. 69–81). Routledge.

Hallmann, K., Rossi, L., Breuer, C., & Ilgner, M. (2020). Determinants of public perception of elite sport. *International Journal of Sport Policy and Politics*, *12*(3), 439–453. https://doi.org/10.1080/19406940.2020.1794930

Hetherington, M. J., & Husser, J. A. (2012). How trust matters: The changing political relevance of political trust. *American Journal of Political Science*, *56*(2), 312–325. https://doi.org/10.1111/j.1540-5907.2011.00548.x

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, *86*(1), 4–29. https://doi.org/10.1162/003465304323023651

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1–14. https://doi.org/10.2307/1269547

Lechner, M., Miquel, R., & Wunsch, C. (2011). Long-run effects of public sector sponsored training in West Germany. *Journal of the European Economic Association*, *9*(4), 742–784. https://doi.org/10.1111/j.1542-4774.2011.01029.x

Lewicki, R. J., & Bunker, B. B. (1995). Trust in relationships: A model of trust development and decline. In B. B. Bunker & J. Z. Rubin (Eds.), *Conflict, cooperation, and justice* (pp. 133–173). Jossey-Bass.

Loomis, J. B. (2011). What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys*, *25*(2), 363–370. https://doi.org/10.1111/j.1467-6419.2010.00675.x

MacDonald, G. M. (1988). The economics of rising stars. *American Economic Review*, *78*(1), 155–166. https://www.jstor.org/stable/1814704

Maennig, W. (2005). Corruption in international sports and sport management: Forms, tendencies, extent and countermeasures. *European Sport Management Quarterly*, *5*(2), 187–225. https://doi.org/10.1080/16184740500188821

Manoli, A. E., Bandura, C., & Downward, P. (2020). Perceptions of integrity in sport: Insights into people's relationship with sport. *International Journal of Sport Policy and Politics*, *12*(2), 207–220. https://doi.org/10.1080/19406940.2020.1747101

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McEvily, B., Perrone, V., & Zaheer, A. (2003). Trust as an organizing principle. *Organization Science*, *14*(1), 91–103. https://doi.org/10.1287/orsc.14.1.91.12814

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, *23*(3), 473–490. https://doi.org/10.5465/amr.1998.926622

Mills, B. M. (2017). Policy changes in Major League Baseball: Improved agent behavior and ancillary productivity outcomes. *Economic Inquiry*, *55*(2), 1104–1118. https://doi.org/10.1111/ecin.12396

Morrison, M., & Brown, T. C. (2009). Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies. *Environmental and Resource Economics*, *44*(3), 307–326. https://doi.org/10.1007/s10640-009-9287-3

Muñiz, C., Rodríguez, P., & Suárez, M. J. (2014). Sports and cultural habits by gender: An application using count data models. *Economic Modelling*, *36*, 288–297. https://doi.org/10.1016/j.econmod.2013.09.053

Nalbantis, G., & Pawlowski, T. (2019). US demand for European soccer telecasts: A between- country test of the uncertainty of outcome hypothesis. *Journal of Sports Economics*, *20*(6), 797–818. https://doi.org/10.1177/1527002518817598

Nalbantis, G., Pawlowski, T., & Coates, D. (2017). The fans' perception of competitive balance and its impact on willingness-to-pay for a single game. *Journal of Sports Economics*, *18*(5), 479–505. https://doi.org/10.1177/1527002515588137

Nüesch, S., & Franck, E. (2009). The role of patriotism in explaining the TV audience of national team games—evidence from four international tournaments. *Journal of Media Economics*, *22* (1), 6–19. https://doi.org/10.1080/08997760902724472

Pawlowski, T. (2013). Testing the uncertainty of outcome hypothesis in European professional football: A stated preference approach. *Journal of Sports Economics*, *14*(4), 341–367. https://doi.org/10.1177/1527002513496011

Pawlowski, T., Nalbantis, G., & Coates, D. (2018). Perceived game uncertainty, suspense and the demand for sport. *Economic Inquiry*, *56*(1), 173–192. https://doi.org/10.1111/ecin.12462

Preston, I., & Szymanski, S. (2003). Cheating in contests. *Oxford Review of Economic Policy*, *19*(4), 612–624. https://doi.org/10.1093/oxrep/19.4.612

Putnam, R. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.

Rodríguez, C., Pérez, L., Puente, V., & Rodríguez, P. (2015). The determinants of television audience for professional cycling: The case of Spain. *Journal of Sports Economics*, *16*(1), 26–58. https://doi.org/10.1177/1527002512471536

Rosen, S. (1981). The economics of superstars. *American Economic Review, 71*(5), 845–858. https://www.jstor.org/stable/1803469

United Nations Educational, Scientific and Cultural Organization (2013, September 19–20). *Conference of parties to the international convention against doping in sport. Fourth session*. Paris, France.

Utz, S., Otto, F., & Pawlowski, T. (2021). "Germany crashes out of world cup": A mixed–method study on the effects of crisis communication on facebook. *Journal of Sport Management, 35*(1), 44–54.

van Doorn, J., & Verhoef, P. C. (2011). Willingness to pay for organic products: Differences between virtue and vice foods. *International Journal of Research in Marketing*, *28*(3), 167– 180. https://doi.org/10.1016/j.ijresmar.2011.02.005

Van Reeth, D. (2013). TV demand for the Tour de France: The importance of stage characteristics versus outcome uncertainty, patriotism, and doping. *International Journal of Sport Finance*, *8*(1), 39–60.

World Anti-Doping Agency. (2019). *World Anti-Doping code 2015 with 2019 amendments*.

Appendix (online appendix)

Online appendix A: Description of study samples and variables

Table A1. Sample representativeness.

	(1) Panel sample	(2) Scenario subsample	(3) Adult German population
Age (in years)	53.10	52.15	50.76
Male	0.57	0.53	0.49
Geographical distribution (state level)			
Baden-Württemberg	0.08	0.06	0.13
Bayern	0.13	0.13	0.16
Berlin	0.05	0.04	0.04
Brandenburg	0.03	0.05	0.03
Bremen	0.01	0.00	0.01
Hamburg	0.03	0.04	0.02
Hessen	0.08	0.07	0.08
Mecklenburg-Vorpommern	0.02	0.02	0.02
Niedersachsen	0.09	0.09	0.10
Nordrhein-Westfalen	0.21	0.19	0.22
Rheinland-Pfalz	0.05	0.06	0.05
Saarland	0.02	0.03	0.01
Sachsen	0.11	0.10	0.05
Sachsen-Anhalt	0.02	0.03	0.03
Schleswig-Holstein	0.04	0.04	0.03
Thüringen	0.03	0.03	0.03
Observations	1,756	271	69,421,785

Notes: Population information is derived from official micro-census data of 2018 from the Federal Statistical Office of Germany (2020).

Reference for Table A1

Federal Statistical Office of Germany (2020, April 8). *Database of the Federal Statistical Office of Germany*. Retrieved April 8, 2020, from https://www-genesis.destatis.de/genesis/online/data?operation=sprachwechsel&language=en

Table A2. Description of variables.

Variables of panel survey	Description	Scale
EAC 2018 TV consumption	Number of competition days watched on TV	Metric
Trust in T&F athletes	Index variable (Cronbach's α = 0.95)	Ordinal
Trusting intention [a)]	"I fully trust most track and field athletes."	Ordinal
Honesty of athletes [a)]	"Most track and field athletes behave honestly."	Ordinal
Fairness of athletes [a)]	"Most track and field athletes behave fairly."	Ordinal
Institution-based trust	Index variable (Cronbach's α = 0.88)	Ordinal
Trust in IAAF	Level of trust that the IAAF ensures a fair competition ("no trust at all" = 1; "complete trust" = 5)	Ordinal
Trust in GAA	Level of trust that the GAA ensures a fair competition ("no trust at all" = 1; "complete trust" = 5)	Ordinal
Generalized trust [a)]	Index variable consisting of three items measuring generalized interpersonal trust (Cronbach's α = 0.78)	Ordinal
Interest in T&F	Interest in watching track and field ("I am not at all interested" = 1; "I am extremely interested" = 5)	Ordinal
Patriotism	Level of pride regarding Germany's achievements in sports ("not proud at all" = 1; "extremely proud" = 5)	Ordinal
Male	Male (yes = 1)	Dummy
Age	Age in years	Metric
Employed	Employed (yes = 1)	Dummy
Education		
Other	Other (yes = 1)	Dummy
Upper secondary school	Upper secondary school (yes = 1)	Dummy
University degree	University degree (yes = 1)	Dummy
Income		
<2,000€	Monthly net household income less than 2,000€ (yes = 1)	Dummy
2,001–4,000€	Monthly net household income between 2,001–4,000€ (yes = 1)	Dummy
>4,001€	Monthly net household income more than 4,001€ (yes = 1)	Dummy
Relationship status		
Married	Married (yes = 1)	Dummy
Single	Single (yes = 1)	Dummy
Living with partner	Single, but living together with partner (yes = 1)	Dummy
Divorced/widowed	Divorced or widowed (yes = 1)	Dummy
Additional variables of scenario	**Description**	**Scale**
Trust outcomes		
Trust in competitors [a)]	"I fully trust the competitors of this 100-meter dash."	Ordinal
Honesty of competitors [a)]	"The competitors of this 100-meter dash behave honestly."	Ordinal
Fairness of competitors [a)]	"The competitors of this 100-meter dash behave fairly."	Ordinal
Trust in no doping [a)]	"I rely on/trust the competitors of this 100-meter dash that they do not take PEDs."	Ordinal
Demand outcomes		
Intention-to-consume	Intention to watch the 100-meter dash live on free TV, if the person has the time (yes = 1)	Dummy
Intention-to-consume (bc)	Bias-corrected intention to watch (certainty to watch the event greater than 70% = 1)	Dummy
Willingness-to-pay	Willingness to pay that the ISTAF event is broadcasted live on public TV (as a share of the public broadcasting fees; 0–210€)	Metric
Treatment condition	Awareness that Justin Gatlin was convicted twice for violating the anti-doping rules (being aware = 1)	Dummy
Expertise in T&F	Knowledge of the current 100m world record; absolute value of the deviation between estimated and actual world record (WR: 9.58s)	Metric

Notes: [a)] Variables are measured on a 5-point Likert-type scale, from "strongly disagree" (1) to "strongly agree" (5).

Online appendix B: Balancing statistics of the matching estimations

Table B1. Balancing statistics with the mean biases for comparison groups in first matching estimation (Model 1).

	Mean values		Mean	Bias	Bias
Variable	Treated	Untreated	difference	(%)	reduction (%)
Interest in T&F	3.43	3.67	-0.24	-20.3	74.0
Expertise in T&F (w2)	0.49	0.71	-0.22	-9.4	90.0
Male	0.70	0.69	0.01	1.6	96.7
Age	54.00	55.50	-1.50	-12.6	11.7
Employed	0.61	0.65	-0.04	-7.4	-112.7
Education (reference: other)					
Upper secondary school	0.17	0.20	-0.03	-5.9	25.0
University degree	0.17	0.11	0.06	17.0	42.0
Income (reference: <2,000€)					
2,001–4,000€	0.63	0.69	-0.06	-12.4	56.2
>4,001€	0.13	0.06	0.07	20.6	-86.2
Relationship (reference: married)					
Single	0.20	0.23	-0.03	-7.7	-51.2
Living with partner	0.20	0.21	-0.01	-3.8	46.2
Divorced/widowed	0.15	0.17	-0.02	-4.1	66.0

Notes: Mean bias of the matched sample is 10.2 (mean bias of unmatched sample is 28.3). "w2" = variable is measured in the second survey wave.

Table B2. Balancing statistics with the mean biases for comparison groups in second matching estimation (Model 2).

	Mean values		Mean	Bias	Bias
Variable	Treated	Untreated	difference	(%)	reduction (%)
Interest in T&F	3.56	3.43	0.13	12.3	83.5
Expertise in T&F (w2)	0.31	0.29	0.02	0.8	99.2
Male	0.68	0.68	0.00	-0.2	99.5
Age	55.00	53.58	1.42	12.8	8.8
Employed	0.59	0.57	0.02	2.9	54.8
Education (reference: other)					
Upper secondary school	0.20	0.21	-0.01	-3.6	-11.6
University degree	0.20	0.18	0.02	4.5	86.8
Income (reference: <2,000€)					
2,001–4,000€	0.66	0.54	0.12	24.2	26.5
>4,001€	0.15	0.15	0.00	0.0	98.3
Relationship (reference: married)					
Single	0.20	0.17	0.03	5.1	23.1
Living with partner	0.17	0.23	-0.06	-15.2	-617.0
Divorced/widowed	0.15	0.21	-0.06	-17.1	-72.3
Trust in T&F athletes	3.62	3.65	-0.03	-3.4	29.7
Institution-based trust	3.21	3.08	0.13	14.6	35.0
Generalized trust	3.33	3.34	-0.01	-0.3	98.7

Notes: Mean bias of the matched sample is 7.8 (mean bias of unmatched sample is 25.2). "w2" = variable is measured in the second survey wave.

Online appendix C: Subsample estimations

Table C1. Marginal effects of the selection models.

Variable	Marginal effects (1) Model 3	(2) Model 4
Interest in T&F	**0.076 (0.022)**	**0.068 (0.050)**
Expertise in T&F (w2)	-0.049 (0.750)	**-0.288 (0.002)**
Male	0.110 (0.222)	0.064 (0.380)
Age	-0.002 (0.575)	-0.002 (0.544)
Employed	-0.037 (0.602)	-0.051 (0.506)
Education (ref.: other)		
Upper secondary school	0.015 (0.856)	0.016 (0.870)
University degree	0.162 (0.125)	0.145 (0.167)
Income (ref.: <2,000€)		
2,001–4,000€	0.096 (0.211)	**0.138 (0.086)**
>4,001€	0.048 (0.680)	0.144 (0.234)
Relationship (ref.: married)		
Single	0.058 (0.538)	0.100 (0.305)
Living with partner	0.073 (0.458)	0.070 (0.479)
Divorced/widowed	0.110 (0.324)	0.181 (0.111)
Trust in T&F athletes [a)]		0.041 (0.463)
Institution-based trust [b)]		-0.039 (0.455)
Generalized trust		0.043 (0.380)
Number of obs./Efron's R^2	208 / 0.168	191 / 0.248

Notes: "w2" = variable is measured in the second survey wave (all other variables are measured in the first wave). The dependent variable in the probit models is a dummy measuring one if respondents are aware of the Gatlin case. Presented are the average marginal effects. *P*-values (in parentheses) obtained from 999 bootstrap replications. Bold numbers indicate significance at conventional levels ($p < 10\%$).

Table C2. Effects of the Gatlin case.

	Model 3 [a)]			Model 4 [b)]		
Outcome variables	ATE	ATET	ATENT	ATE	ATET	ATENT
Trust in competitors	**-0.784**	**-0.915**	**-0.748**	**-0.861**	**-0.817**	**-0.883**
	(0.000)	**(0.005)**	**(0.001)**	**(0.001)**	**(0.013)**	**(0.009)**
Honesty of competitors	**-0.625**	**-0.927**	**-0.540**	**-0.891**	**-0.887**	**-0.893**
	(0.003)	**(0.002)**	**(0.016)**	**(0.000)**	**(0.008)**	**(0.012)**
Fairness of competitors	-0.129	**-0.815**	0.064	**-0.735**	**-0.758**	**-0.724**
	(0.730)	**(0.014)**	(0.903)	**(0.008)**	**(0.016)**	**(0.032)**
Trust in no doping	**-0.561**	**-1.083**	-0.414	**-0.986**	**-0.904**	**-1.026**
	(0.017)	**(0.003)**	(0.136)	**(0.008)**	**(0.013)**	**(0.027)**
Intention-to-consume	-0.040	0.065	-0.069	-0.111	-0.059	-0.137
	(0.644)	(0.766)	(0.414)	(0.272)	(0.653)	(0.258)
Intention-to-consume (bc)	0.039	0.037	0.039	-0.027	0.033	-0.056
	(0.730)	(0.823)	(0.703)	(0.785)	(0.887)	(0.598)
Willingness-to-pay	24.546	-9.044	33.986	-9.493	-3.357	-12.524
	(0.279)	(0.370)	(0.163)	(0.195)	(0.752)	(0.145)

Notes: Presented are the treatment effects (ATE, ATET, & ATENT) of being aware of the Gatlin case. [a)] number of observations = 208, common support = 196 (94%), number of treated = 43; [b)] number of observations = 191, common support = 121 (63%), number of treated = 40. *P*-values (in parentheses) obtained from 999 bootstrap replications. Bold numbers indicate significance at conventional levels ($p < 10\%$).

4.2 Political relations and sports: Exploring the demand for relocated soccer games*

EUROPEAN SPORT MANAGEMENT QUARTERLY
https://doi.org/10.1080/16184742.2022.2092171

Routledge
Taylor & Francis Group

Check for updates

Political relations and sports: exploring the demand for relocated soccer games

Felix Otto [a], Georgios Nalbantis [a] and Tim Pawlowski [a,b,c]

[a]Faculty of Economics and Social Sciences, Institute of Sports Science, University of Tübingen, Tübingen, Germany; [b]Interfaculty Research Institute for Sports and Physical Activity, University of Tübingen, Tübingen, Germany; [c]LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany

ABSTRACT

Research question: In recent years, European soccer leagues have strengthened their internationalization efforts by relocating games abroad. In this context, we empirically investigate the impact of the political relations between the importing and the exporting countries on on-site demand for such games. While there is a rich literature addressing this issue in more general settings, it has never been explored in sports.

Research methods: We use a stated-preference approach and survey data of a representative sample of soccer-interested US residents, which we combine with data from the American Community Survey. The effects are estimated using Tobit regressions and an instrumental variable approach.

Results and findings: We find that the willingness-to-pay for admission to relocated games is higher if the country where the competing teams come from is perceived as friendly.

Implications: We provide preliminary empirical evidence that the political relations between importing and exporting countries may be important for international demand of spectator sports. As such, the public perception about political relations between countries may constitute a crucial component for sport leagues' expansion strategies.

ARTICLE HISTORY
Received 12 October 2021
Accepted 14 June 2022

KEYWORDS
Country of origin; international political relations; internationalization strategies; soccer leagues; spectator demand; instrumental variable

* Acknowledgements: Previous versions of the paper were presented at the 24th Annual Meeting of the German Association of Sport Economics and Sport Management in 2021; the 2nd North American Association of Sports Economists (NAASE) Virtual Conference in 2021; the 12th Conference of the European Sport Economics Association in 2021. We thank NAASE discussant Brad Humphreys and conference participants for their valuable comments and suggestions. The usual disclaimer applies.

Funding: The authors disclose receipt of the following financial support for the research, authorship, and/or publication of this article: Data collection was partly supported by a FIFA/CIES Joao Havelange Scholarship, awarded to Georgios Nalbantis and Tim Pawlowski. The content of this article reflects only the authors' views, and the funders are not liable for any use that may be made of the information contained therein. The funders had no role in the study design, data collection and analysis, or preparation of this research. Any remaining errors are our own.

Political relations and sports: Exploring the demand for relocated soccer games

Introduction

Professional soccer leagues in Europe are commonly organized as domestic contests. However, as domestic markets become increasingly saturated, there is a continuing need for leagues to generate additional revenues in order to remain competitive, both on and off the field. This has recently led several leagues to strengthen their expansion efforts to international markets. A strategy in this regard is the relocation of games to foreign countries. Most commonly is the relocation of so-called super cups, which are usually single games between the recent winner of the national cup and the domestic league champion just before the start of the new season. For instance, the French super cup has been staged in eight different foreign countries in the last 13 years. The Italian super cup was staged seven times abroad in Saudi Arabia, China, and Qatar in the last 10 years, while the first relocated editions of the competition were hosted twice in the US. The Spanish super cup has been relocated to Saudi Arabia since 2019 and is now staged as a four-team contest. Apart from that, there have also been ongoing discussions among league officials in the English Premier League, the Italian Serie A, and the Spanish La Liga about staging regular league games abroad.[1] This is already common practice in North American major leagues like the National Football League (e.g. NFL International Series in London and Mexico City; Munich and Frankfurt will host league games from the 2022 season onwards) and Major League Baseball (e.g. regular season games in London, Tokyo, and Monterrey).

Even if severe political conflicts are absent, the economic success of such strategies may depend on the political relations between the importing and the exporting country; that is, the country of origin (COO). Theoretically, depending on whether the COO is perceived as friendly to the importing country (or not), countries' similarities (or differences)

[1] For a media discussion on relocating games of La Liga to the US see: https://www.theguardian.com/football/2018/aug/16/la-liga-stage-match-us. For Serie A discussions to relocate games to China see: https://www.reuters.com/article/uk-soccer-italy-china/italy-may-discuss-staging-serie-a-match-in-china-idUKKCN1R320B. For Premier League discussions adding a 39th game in the calendar to be played abroad see: https://www.dailymail.co.uk/sport/football/article-5207145/Manchester-City-plans-39th-Premier-League-game.html.

become salient and sport consumers may use their national identity for in- and out-group formation, affecting their individual preferences and demand (Akerlof & Kranton, 2000). Unlike many other products that just use the COO as a simple marketing cue (e.g. the "Made in" label), national sport leagues are inevitably associated with their COO. As such, according to social identity theory (Tajfel & Turner, 1986) and its feature of in-group favoritism (e.g. Rand et al., 2009), political relations between both countries may alter the demand for relocated games.

Despite the increasing relevance of internationalization processes in sports in general and the potential impact of countries' political relations on spectator sports demand in particular, this topic has so far not been empirically explored in the literature. The few existing studies on international sports demand have either focused on fan preferences for game uncertainty (Schreyer et al., 2018; Nalbantis & Pawlowski, 2019) or substitution effects (Nalbantis et al., 2021). Other related studies have focused on the role of patriotism for the demand of televised sports, finding that individuals prefer to watch their own national teams (Bennett et al., 2007; Nuesch & Franck, 2009) or compatriot athletes in individual sports (Kim et al., 2008; Konjer et al., 2017; Van Reeth, 2013). With regard to relocations of US sport *teams,* contingent valuation method (CVM) studies frequently found that the costs of relocations exceed the value of public goods such as civic pride and sense of community (Johnson et al., 2001; Johnson et al., 2007; Johnson & Whitehead, 2000). However, little is known about relocations of *single games* in international settings and the role of political relations between countries when sports are considered *imported goods*.

The literature exploring such issues in other settings is rich (for a review, see Riefler & Diamantopoulos, 2007). A common finding is that international political conflicts and ensuing boycott calls negatively impact US consumer demand for products of the exporting country (e.g. Heilmann, 2016; Pandya & Venkatesan, 2016). Moreover, consumer animosity toward countries negatively impacts product evaluation and demand, regardless of the product quality (e.g. Klein et al., 1998). However, since these issues have typically been explored in the context of historic or current crises between two countries (e.g. Leong et al., 2008), little is known about the role of political relations in the absence of severe conflicts (Funk et al., 2010). Moreover, most studies have focused on durable goods such as cars or electronic devices. However, consumers who avoid

brands from less favorable countries are expected to sacrifice less if products are non-durable and easy to substitute – like spectator sports events (Riefler & Diamantopoulos, 2007).

In this paper, we examine whether the perceived political relations between countries may impact spectator demand for relocated soccer games. We focus on the so-called "Big Five" European leagues, with England (Premier League), Spain (La Liga), France (Ligue 1), Italy (Serie A), and Germany (Bundesliga) as exporting countries and the US as the importing country. This setting seems promising for various reasons. *First*, the development of the soccer market in the US has been highly dynamic, offering great potential for European leagues (Nalbantis & Pawlowski, 2016). In fact, the US market already contributes significantly to the international media revenues generated by these leagues (Nalbantis et al., 2021). *Second*, product ethnicity for European leagues is generally higher than it is for the domestic Major League Soccer (MLS) (Szymanski, 2003).[2] *Third*, the US consists of many regional markets, with traditionally distinct political attitudes between residents (Abramowitz & Saunders, 2008); this enables us to exploit some regional variation for identification in our estimation approach.

In our analysis, we use survey data of a representative sample of soccer-interested US residents and explore the relation between perceived friendliness toward the COO and willingness-to-pay (WTP) for (hypothetical) admission to relocated European soccer games. In an attempt to address endogeneity concerns, we also implement an instrumental variable (IV) approach using the share of veterans relative to the adult population per county as IV. In line with the theory, our findings suggest that US consumers have a higher WTP for (hypothetical) admission to relocated European soccer games if the COO is perceived as friendly to the US in international affairs. As such, we provide preliminary empirical evidence that countries' political relations, which are unrelated to the product's quality or supply, may impact the international demand for spectator sports.

The remainder of this paper is organized as follows. The next section describes the data and variables used, followed by a discussion of our empirical strategy. We then present the results and the final section concludes.

[2] Product ethnicity describes the stereotypical associations of a product with particular countries: for example, red wine is typically associated with Italy and France (Usunier & Cestre, 2007).

Data

The relation of interest is tested with a stated-preference approach for hypothetically relocated cup finals as well as league games. We consider this approach as reasonable because identifying the impact of political relations seems to be challenging with higher aggregated data (e.g. sales data; see Chavis & Leslie, 2009; Pandya & Venkatesan, 2016) given the idiosyncratic nature of political attitudes. We consider domestic cup finals as an ideal setting to examine the role of political relations for spectator demand because these contests have an explicit national character, and the COO is salient for the spectators. For instance, national anthems are typically played before finals, while state representatives, such as prime ministers, presidents, or members of the royal family, present the trophy to the winning team. Moreover, the quality of the competing teams is usually high in cup finals, ensuring considerable fan interest. In order to test whether the type of game matters, we also consider league games as a second case.

Data collection

Survey participants were randomly recruited from a US-wide representative online panel.[3] The survey contained a comprehensive set of questions for approximating consumer preferences and demand for professional soccer in the US. The first survey was conducted online in May 2015 before the national cup finals of four major European soccer leagues: i.e., the English FA Cup, the Coup de France, the German DFB-Pokal, and the Spanish Copa del Rey.[4] The second survey was conducted online at the beginning of September 2015 and focused on games from the top five European leagues: i.e., the English Premier League, the French Ligue 1, the German Bundesliga, the Italian Serie A, and the Spanish La Liga. The selection of league games was based on quality and popularity of the contestants. Table 1 presents an overview of all games included in the two surveys.

[3] The survey was administered by the globally operating market research company Questback. The online panel was provided by Lightspeed GMI.

[4] Since the Italian cup final (Coppa Italia) was rescheduled on short notice and staged during our survey period, it was excluded on purpose from the cup finals sample ex post. To check whether the absence of that game and country affects the findings in any form, (i) we also excluded the corresponding Italian Serie A game from the league games sample. Likewise, we checked, (ii) whether including the Italian cup final in the cup finals sample affects our main findings. Since our main conclusions hold for both re-estimations, we are not concerned about any bias due to omitting the Italian cup final in our main specification (results are available upon request).

Table 1. Overview of the soccer games included in the surveys.

Cup finals	Country	Team A	Team B	Date
FA Cup	England	Arsenal FC	Aston Villa FC	30 May 2015
Coup de France	France	AJ Auxerre	Paris St.-Germain FC	30 May 2015
DFB-Pokal	Germany	BVB Dortmund	VfL Wolfsburg	30 May 2015
Copa del Rey	Spain	Athletic Club de Bilbao	FC Barcelona	30 May 2015
League games	Country	Team A	Team B	Date
Premier League	England	Everton FC	Chelsea FC	12 Sep. 2015
Ligue 1	France	Paris St.-Germain FC	FC Girondins de Bordeaux	12 Sep. 2015
1. Bundesliga	Germany	FC Bayern Munich	FC Augsburg	12 Sep. 2015
Serie A	Italy	FC Internazionale Milano	AC Milan	13 Sep. 2015
La Liga	Spain	Club Atlético de Madrid	FC Barcelona	12 Sep. 2015

A screen-out question and a quality corrections measure were used to exclude respondents with no interest in soccer and low response quality, respectively. After also excluding observations with missing values, we obtained information from 2645 and 2619 respondents from the first and second survey, respectively.[5] As respondents provide information on each of the games in the respective survey, our sample consists of 10,580 and 13,095 game-level observations for cup finals and league games, respectively.

Willingness-to-pay measure

Similar to other analyses of spectator sports demand (e.g. Nalbantis et al., 2017), we employ a WTP measure as demand proxy. Our outcome variable measures the respondents' WTP for admission to the aforementioned games in the hypothetical scenario that these games would be staged in a stadium close to their residence at a time and date that would be convenient to attend. This scenario allows us to control for potential scheduling issues and opportunity costs such as travel costs. We use an open-ended elicitation format to avoid potential *anchoring bias* (Green et al., 1998).

Figure 1 displays the distribution of our WTP measure. It shows that more than 50% of respondents stated zero WTP for both the cup finals and league games. Zero values can indicate that individuals are not willing to pay any money for a game or, alternatively, it can be some form of protest. However, in contrast to CVM and the valuation of non-

[5] See Nalbantis and Pawlowski (2016) for a detailed description of the survey design, data quality, and representativeness of the samples (e.g. geographical distribution). In total, 6590 persons were invited to participate in the first survey and the response rate was 47%. In total, 5805 persons were invited for the second survey and the response rate was 54%. In total, 44% of respondents from the second survey also participated in the first survey. Moreover, our survey includes US residents, i.e., both US and non-US citizens. Excluding those with no US citizenship (< 4% of the sample) does not affect our findings (results are available upon request).

marketed goods (e.g. see Lo & Jim, 2015), the presence of *protest zeros* is unlikely in our scenario because our elicitation is based on the evaluation of a common marketed good (i.e. tickets), a reasonably uncontroversial setting, and because organizers are not expected to offer free admission to soccer games. Another source of zeros could be generated by respondents who protest to issues related to the questionnaire (e.g. length, topic, design, etc.) or simply because zero is perceived as an "easier answer" for those with small valuations (e.g. see Castellanos et al., 2011). As these two types of zeros relate to the ease of answering the questionnaire, we argue that the occurrence of such types of zeros is positively correlated with the speed of answering the questionnaire. Since we made use of quality checks and removed survey speedsters, we are confident that the likelihood of having any such zeros in our estimation samples is considerably reduced. Accordingly, we interpret zero WTP statements as true expressions of choice for not being willing to pay any money, and thus, no demand for soccer games.

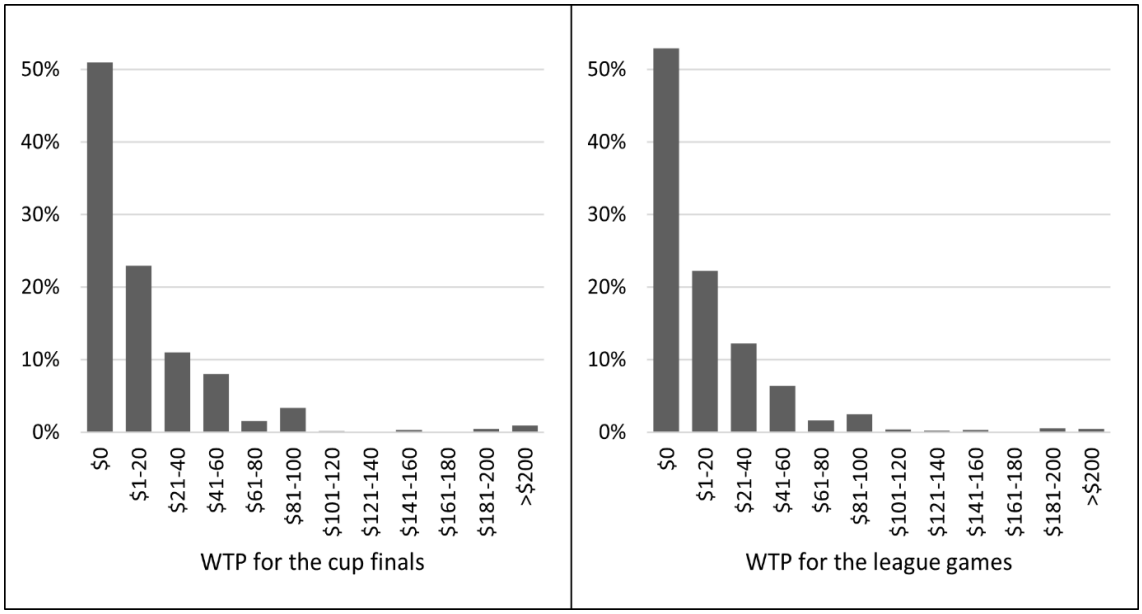

Figure 1. Distribution of the unwinsorized willingness-to-pay (WTP).

Notes: The data stem from two surveys on cup finals and league games; number of observations: 10,580 in the cup finals sample and 13,095 in the league games sample; the WTP-variable measures the respondents' maximum WTP for a ticket to attend the games if they would be relocated and take place in a stadium near the respondent's residence. In this figure, the WTP statements are collapsed into categories (in intervals of $20) for illustrative purposes.

In order to mitigate any *strategic bias*, respondents were informed that the research project would only be used for academic purposes. This should have raised awareness that their statements could not have any (in)direct impact on the pricing policy of the

leagues or teams. Finally, we expect that the soccer-interested sample consists of respondents who are familiar with the product of choice – that is, the purchase of soccer tickets – which should mitigate *hypothetical bias*, as shown by Schläpfer and Fischhoff (2012). In fact, the distribution of WTP shows that a large share of WTP statements is under $100 (see Figure 1), corresponding to the typical price segment for soccer tickets in the US (see Nalbantis & Pawlowski, 2016). Apart from that, as shown by Miller et al. (2011), even if a hypothetical bias is present, the open-ended method can still accurately predict demand. Moreover, we decided to winsorize the *WTP* variable at the 99th percentile in order to reduce the potential impact of outliers in the analysis.[6] The average winsorized and unwinsorized *WTP* for attending relocated cup finals is $18 and $20, respectively; the average winsorized and unwinsorized *WTP* for attending league games is around $17, respectively (see Table 2).

Political relations measure

The respondents' perception of the countries' political relations is assessed with a proxy that is commonly used in COO research, i.e., the perceived friendliness toward the respective countries in international affairs (Parameswaran & Yaprak, 1987), measured on an ordinal five-point Likert scale. Figure 2 displays the distribution of the perceived *country friendliness* towards the different countries of origin in both surveys. For each country under consideration, we observe considerable heterogeneity between *individuals*, with most respondents indicating a positive perception (ratings 4 and 5). At the same time, the average ratings across the different *countries* reveal a fairly similar pattern, except for the United Kingdom (UK), which is regarded as the friendliest country among the five countries under consideration.[7] This is in line with results frequently reported in US opinion polls (e.g. Loschky & Riffkin, 2015). Moreover, by comparing the results between the first and second surveys we observe a remarkable stability over time with regard to the responses averaged across respondents. This also supports the choice of our

[6] The value corresponding to the 99th percentile is assigned to all values above this threshold, which is the case for 98 and 131 observations in the cup finals and league games samples, respectively. This approach (also called winsorizing) reduces the impact of outliers, while keeping all observations in the analysis (Ghosh & Vogt, 2012). We prefer this method as larger stated values usually contain more hypothetical bias (Murphy et al., 2005). Outlier reduction via trimming produces similar results (results are available upon request).

[7] The rather small variation of perceived friendliness *between* countries (see Figure 2) may suggest that our findings are only driven by the different perceptions of the UK. We therefore re-estimate our main models excluding the UK games (i.e., the FA Cup final and Premier League game). Our findings remain (results are available upon request).

IV, which will be explained further below. As indicated by ratings 4 and 5, about 66% of respondents in both samples perceive the countries of origin as friendly (see Table 2).

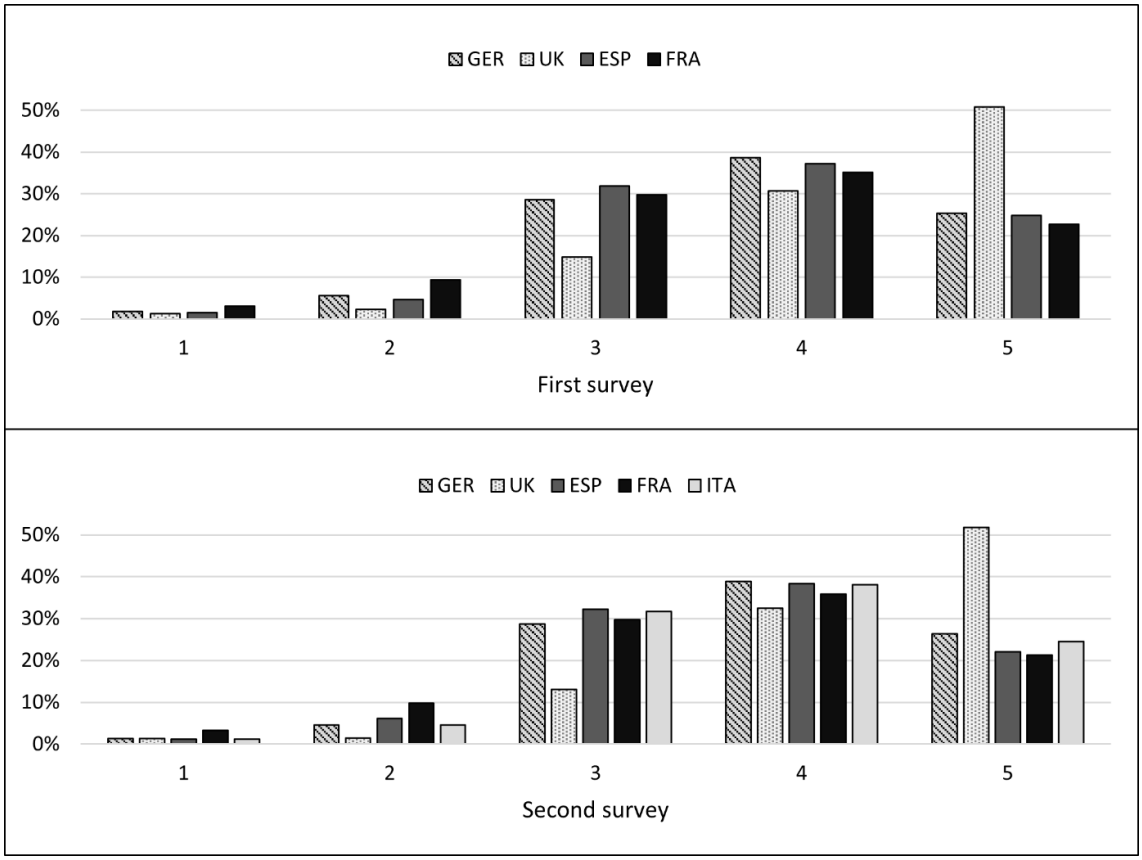

Figure 2. Distribution of the perceived friendliness toward the countries of origin.

Notes: The data stem from two surveys on cup finals (first survey) and league games (second survey); number of observations: 10,580 in the cup finals sample and 13,095 in the league games sample; the variable measures the respondents' perception if the following countries are friendly to the US in international affairs, on a five-point Likert scale, ranging from 1=strongly disagree to 5= strongly agree. Abbreviations: GER=Germany, UK=United Kingdom, ESP=Spain, FRA=France, ITA=Italy.

Control variables

In order to establish the relationship between *country friendliness* and *WTP*, we control whether respondents are supporters of any of the clubs in the respective leagues and of the competing clubs in the respective games. Moreover, we control for game uncertainty using a measure of the individually perceived winning probabilities of the teams and its squared term, as suggested by Pawlowski et al. (2018). Although our hypothetical scenario approach allows us to mitigate potential opportunity costs, respondents may have a local (soccer) stadium experience in mind when assessing and stating their WTP. To proxy this, we control whether an MLS stadium is located within a 10-mile radius around

the respondents' residence. Finally, we control for several socio-demographic variables that were previously found to be related to WTP for soccer tickets (e.g. Nalbantis et al., 2017).[8] Table 2 presents all variables included in the estimations as well as their descriptive statistics.

Overall, we observe marginal differences between both samples with regard to the descriptive statistics, partly due to an overlap of respondents in both surveys (see footnote 5). In the cup finals sample, around 3% of respondents are supporters of one of the contestants in the finals, while around 14% support another club from the respective countries. In total, 11% are Hispanophone, 81% are whites, 63% are married, 53% are male, the average age is 48 years, and they live on average in households with 2.7 members. In the league games sample, about 4% of respondents are supporters of one of the competing clubs and 11% are supporters of other clubs from the respective leagues. In total, 10% are Hispanophone, 86% are whites, 65% are married, 56% are male, the average age is 51 years, and they line on average in households with 2.6 members. In both samples, around 10% of respondents live within a 10-mile radius of an MLS stadium, 6% are unemployed, more than 50% have a college degree, and above 60% earn more than $50,000 per year.

[8] To check whether our findings differ by geography, we run further models including state dummies. Our main findings remain. Moreover, state dummies are jointly statistically insignificant (results are available upon request).

Table 2. Description of the variables and descriptive statistics.

			Cup finals			League games		
Variable name	Description	Scale	Mean	SD	Min-Max	Mean	SD	Min-Max
WTP	Maximum willingness-to-pay for a ticket to attend the respective game if it would be relocated and take place in a stadium near the respondent's residence at a time and date that would be convenient to attend (in US dollars). Extreme values are winsorized at the 99th percentile.	M	18.200	33.432	0–200	16.502	31.171	0–199
Country friendliness 1	COO friendliness to USA in international affairs, rating level 1 (yes = 1)	D	0.019	0.138	0–1	0.017	0.128	0–1
Country friendliness 2	COO friendliness to USA in international affairs, rating level 2 (yes = 1)	D	0.055	0.228	0–1	0.053	0.225	0–1
Country friendliness 3	COO friendliness to USA in international affairs, rating level 3 (yes = 1)	D	0.262	0.440	0–1	0.271	0.444	0–1
Country friendliness 4	COO friendliness to USA in international affairs, rating level 4 (yes = 1)	D	0.354	0.478	0–1	0.367	0.482	0–1
Country friendliness 5	COO friendliness to USA in international affairs, rating level 5 (yes = 1)	D	0.309	0.462	0–1	0.292	0.455	0–1
Club supporter (game)	Being a supporter of a club in the respective game (yes = 1)	D	0.028	0.164	0–1	0.036	0.187	0–1
Club supporter (league)	Being a supporter of another club in the respective league (yes = 1)	D	0.139	0.346	0–1	0.107	0.309	0–1
Perceived team A win Pr	Subjective evaluation of the home teams' (team A) winning probability (team A will definitely win= 0; team B will definitely win = 10)	M	4.578	2.474	0–10	4.638	2.389	0–10
Nearby MLS stadium	MLS stadium is located in a 10-mile radius around residence (yes = 1)	D	0.103	0.303	0–1	0.097	0.296	0–1
Hispanophone	Being Hispanic, Latino, or Spanish origin (yes = 1)	D	0.113	0.317	0–1	0.103	0.304	0–1
White	Being white (yes = 1)	D	0.805	0.396	0–1	0.857	0.350	0–1
Male	Being male (yes = 1)	D	0.528	0.499	0–1	0.559	0.496	0–1
Married	Being married (yes = 1)	D	0.630	0.483	0–1	0.646	0.478	0–1
Household size	Number of household members	M	2.745	1.361	1–7	2.640	1.308	1–7
Age (in years)	Age in years	M	47.987	13.979	18–83	50.797	14.712	17–96
Unemployed	Being currently unemployed (yes = 1)	D	0.062	0.241	0–1	0.057	0.231	0–1
No high school	No high school diploma (yes = 1)	D	0.005	0.073	0–1	0.003	0.055	0–1
High school diploma	High school diploma (yes = 1)	D	0.128	0.334	0–1	0.134	0.341	0–1
Some college	Attended college but no degree or associate degree (yes = 1)	D	0.339	0.473	0–1	0.352	0.478	0–1
Bachelor's degree	Bachelor's degree (yes = 1)	D	0.358	0.479	0–1	0.330	0.470	0–1
Graduate degree	Master's, doctoral, or professional degree (yes = 1)	D	0.170	0.376	0–1	0.181	0.385	0–1
Income <$25K	Yearly net household income less than <$24.99K (yes = 1)	D	0.142	0.349	0–1	0.130	0.337	0–1
Income $25–49.99K	Yearly net household income between $25–49.99K (yes = 1)	D	0.242	0.428	0–1	0.232	0.422	0–1
Income $50–74.99K	Yearly net household income between $50–74.99K (yes = 1)	D	0.229	0.421	0–1	0.243	0.429	0–1
Income $75–100K	Yearly net household income between $75–99.99K (yes = 1)	D	0.154	0.361	0–1	0.157	0.364	0–1
Income >$100K	Yearly net household income more than >$100K (yes = 1)	D	0.233	0.423	0–1	0.238	0.426	0–1
IV veteran share (ACS)	Share of veterans relative to the adult population per county	M	0.085	0.031	0.025–0.269	0.086	0.032	0.025–0.269

Note: The data stem from two surveys on cup finals (first survey) and league games (second survey); number of observations: 10,580 in the cup finals sample and 13,095 in the league games sample. The WTP variable is winsorized at the 99th percentile. Data for *IV veteran share* stems from the American Community Survey (ACS). The 10-mile radius for *Nearby MLS stadium* was determined using geographic coordinates of the survey respondents and the closest MLS stadium. All other variable information comes from respondents' statements in the online surveys. COO=country of origin; M=metric variable; D=dummy variable; SD=standard deviation.

Empirical strategy

We start our analysis by regressing our *WTP* measures for cup and league games on *country friendliness* and the control variables. Since our outcome variables are left-censored, including a large amount of zero statements, we employ Tobit regressions, estimating the following model:

$$WTP_{ig} = b_1 country\ friendliness_{ig} + b_2 X_{1_{ig}} + b_3 X_{2_i} + \mu_g + \varepsilon_{1_{ig}} \quad (1)$$

where WTP_{ig} is the WTP of individual *i* for admission to relocated game *g*. The variable $country\ friendliness_{ig}$ denotes individual *i*'s friendliness rating for the COO of game *g*. $X_{1_{ig}}$ is our set of individual-specific control variables for the games (e.g. supporter status and perceived game uncertainty). X_{2_i} is our set of individual-specific socio-demographic control variables (e.g. age, ethnic background, or income). In order to control for unobserved differences between the games (e.g. popularity of the competing teams or (star) players) and countries (e.g. language similarities or shared history), we include *game/country dummies* (μ_g). Finally, as the data on the different games stem from the same respondents, we cluster standard errors at the individual level.

In the initial approach, however, endogeneity concerns arise because data for our outcome and explanatory variables were gathered from the same respondents with the same instrument. This may lead to common method variance and reverse causality issues (Antonakis et al., 2010). Following Sande and Ghosh (2018), we attempt to address these concerns, using an IV approach with an exogenous variable that correlates with *country friendliness* without having an independent effect on *WTP*. Our idea is to proxy the level of national identity in the respondents' area of residence. National identity has been frequently found to negatively influence the attitude toward foreign countries (e.g. Shoham et al., 2006) and we argue that it should influence the *WTP* for relocated games only through the respondent's perceived friendliness toward the COO.

To proxy national identity of the respondents, we consider the share of veterans relative to the adult population per county, which is determined outside of our survey sample. This data stems from the 2015 American Community Survey (ACS) provided by the United States Census Bureau (2020). We pooled the ACS data with our survey data by

taking advantage of the respondents' ZIP code statements.[9] We argue that this is a suitable instrument because: *first*, military service is frequently argued to stimulate the formation of national identity (Cáceres-Delpiano et al., 2021); *second*, veterans are more likely to live in areas where national identification is very salient (e.g. due to military bases, see Teachman, 2013); *third*, veterans may shape political attitudes of their social environment (Goetz, et al., 2019; Krueger & Pedraza, 2012). We thus expect respondents living in regions characterized by a strong national identity (i.e. a high *veteran share*) to have less favorable attitudes toward foreign countries. In other words, there should be a negative correlation between *IV veteran share* and *country friendliness* across all countries under consideration.

In order to satisfy the exclusion restriction, the choice of this IV relies on the following underlying assumptions. *First*, although we have no information on the veteran or military status of our survey respondents, we have no reason to believe that we have many veterans in our survey sample, which could bias our results. *Second*, it is rather unlikely that veterans' individual tastes or preferences directly affect the WTP of our survey respondents. For example, if a respondent is living in a neighborhood with some veterans, one cannot expect that she knows the veterans' tastes for European soccer. Moreover, even if this would be the case, her *WTP* is unlikely to be affected by the veterans' tastes. At least, such spillover effects should be small and may only exists for a small number of respondents. Thus, a direct effect of *veteran share* on *WTP* should be of less concern. *Third*, the *veteran share* could affect the respondents' *WTP* through channels other than national identity. For example, counties with more veterans could have higher income levels or less unemployment; however, such effects should be captured by the inclusion of socio-demographic variables as controls in our models. Higher or lower *veteran shares* may also be related to differences in lifestyle characteristics of the local population that we do not control. Although we are not aware of such potential channels, we acknowledge that this could eventually violate the exclusion restriction. Therefore, while being

[9] The survey respondents were assigned a value that indicates the share of veterans in the county they live in according to their ZIP code statements. To identify the county for each ZIP code, we use data from https://www.unitedstateszipcodes.org/. In case ZIP codes cross county borders, the website reports the most common county considering the percentage of population of a ZIP code area in the county areas as well as the percentage of the county areas in the ZIP code.

confident that our IV is valid, we remain cautious and provide these estimations as an additional effort to test the relation of interest.

Since our *WTP* measure is censored at zero and our *country friendliness* measure is ordinal, we rely on Roodman's (2011) conditional (recursive) mixed process (CMP) estimator, estimating the following equations:

$$WTP_{ig} = b_1 country\ \widehat{friendliness}_{ig} + b_2 X_{1_{ig}} + b_3 X_{2_i} + \mu_g + \varepsilon_{2_{ig}} \quad (2)$$

$$country\ friendliness_{ig} = \alpha_1 IV\ veteran\ share_{ci} + b_2 X_{1_{ig}} + b_3 X_{2_i} + \mu_g + \varepsilon_{3_{ig}} \quad (3)$$

where the instrument is the *IV veteran share* in county *c* of individual *i*. The variable $country\ \widehat{friendliness}_{ig}$ in equation (2) is the predicted value of *country friendliness* from the first-stage regression in equation (3). In the IV case, CMP implements a limited-information maximum likelihood estimator, which allows to estimate an ordered-probit regression in the first stage if an ordinal endogenous variable appears in the second stage of the IV-Tobit model. The relevance of our IV is supported by our first-stage results, which indicate a negative and statistically significant relationship between *veteran share* and *country friendliness* (see Table A1 in Online Appendix A).[10]

Results

Following McDonald and Moffitt (1980), we report the unconditional marginal effects on the expected value of *WTP* for all cases (including zero and positive values) of the Tobit estimations in Table 3. We begin by discussing the results for our political relations measure by comparing the unfriendly ratings (country friendliness 1 and 2) and friendly ratings (country friendliness 4 and 5) to country friendliness 3, which we interpret as a neutral category where persons might be either indifferent or ambivalent.

[10] A caveat of the CMP estimator is that the software package lacks common tests to further assess the instrument strength. We try to circumvent this limitation with the use of two-stage least squares (2SLS) estimations. While this approach yields less efficient estimates, as it does not consider the limited nature of our data, it can provide commonly accepted diagnostic tests, the results of which can be seen as indicative (Angrist & Pischke, 2009). We consulted the effective F-statistic and the weak-instrument-robust Anderson-Rubin confidence set, which suggest that when using a linear model, the IV is only valid for the cup finals estimations. Therefore, we are more cautious when interpreting results for the league games sample.

Political relations measure

According to our Tobit models, we find a statistically significant and positive relation between *country friendliness* and the *WTP* for both types of games. In detail, the expected *WTP* is \$6 higher for cup finals and \$5 higher for league games if the COO is perceived as very friendly compared to a neutral perception. In contrast, the estimates for all unfriendly ratings are comparably less precise. The second-stage results of our IV-Tobit models confirm the significant relation between *country friendliness* and *WTP*, including statistically significant differences between a neutral perception and both friendly and unfriendly perceptions. The expected *WTP* is \$4 lower for cup finals and \$3 lower for league games if the COO is perceived as rather unfriendly. If the COO is perceived as rather friendly, the expected *WTP* is \$12 higher for cup finals and \$8 higher for league games. If the COO is perceived as very friendly, the expected *WTP* is even \$31 higher for cup finals and \$18 higher for league games. Overall, the IV-approach suggests that the coefficients of *country friendliness* are downward-biased when following the initial approach, which is quite common when facing endogeneity issues in cross-sectional survey data (Antonakis et al., 2010; Becker, 2016). Moreover, a comparison between the Tobit and IV-Tobit estimates shows that differences between ratings are more pronounced in the IV approach. Additionally, in the IV approach, the effects are constantly growing with scale. This could indeed suggest a measurement error caused by common method variance in the political relation measure.

Table 4 reports the predicted average *WTP* based on the Tobit and IV-Tobit estimates. According to the Tobit estimates, the predicted average *WTP* is \$20 for cup finals and \$18 for league games. When the COO is perceived as very friendly, the predicted average *WTP* is \$22 for cup finals and \$20 for league games. According to the IV Tobit estimates, the predicted average *WTP* is \$24 for cup finals and \$19 for league games. When the COO is perceived as very friendly, the predicted average *WTP* is \$41 for cup finals and \$29 for league games. The estimated differences in the magnitude of the effect of interest support the notion that cup finals may have a more salient COO.

Table 3. Results for the willingness-to-pay for relocated soccer games.

	Cup finals		League games	
	Tobit	IV-Tobit	Tobit	IV-Tobit
Country friendliness 1	-4.025* (2.187)	-8.334*** (0.726)	1.231 (3.190)	-4.971** (1.965)
Country friendliness 2	1.864 (1.390)	-4.005*** (0.769)	-0.178 (1.091)	-3.465*** (1.008)
Country friendliness 3	Reference cat.	Reference cat.	Reference cat.	Reference cat.
Country friendliness 4	3.511*** (0.818)	11.616*** (1.311)	3.206*** (0.764)	7.884*** (1.648)
Country friendliness 5	5.685*** (1.034)	30.973*** (4.656)	4.839*** (0.954)	18.093*** (5.280)
Club supporter (game)	18.130*** (2.442)	14.718*** (2.607)	17.446*** (1.947)	15.359*** (2.032)
Club supporter (league)	14.448*** (1.223)	10.475*** (1.257)	15.846*** (1.434)	13.359*** (1.566)
Perceived team A win Pr	-5.614*** (0.527)	-4.086*** (0.602)	-4.297*** (0.470)	-3.325*** (0.594)
Perceived team A win Pr^2	0.506*** (0.052)	0.378*** (0.058)	0.380*** (0.047)	0.296*** (0.056)
Nearby MLS stadium	-0.775 (1.584)	-1.028 (1.719)	1.470 (1.446)	1.420 (1.508)
Hispanophone	4.261** (1.683)	3.250* (1.746)	5.837*** (1.678)	5.153*** (1.736)
White	-5.753*** (1.284)	-5.900*** (1.355)	-1.822 (1.196)	-2.124* (1.240)
Male	2.450*** (0.941)	1.779* (1.029)	2.451*** (0.874)	2.032** (0.913)
Married	-0.898 (1.114)	-0.783 (1.212)	-0.439 (1.074)	-0.113 (1.104)
Household size	0.678* (0.406)	0.734 (0.452)	1.551*** (0.470)	1.465*** (0.483)
Age (in years)	-0.408*** (0.037)	-0.455*** (0.042)	-0.339*** (0.038)	-0.365*** (0.040)
Unemployed	-2.744 (1.755)	-2.655 (1.966)	-0.898 (1.865)	-1.154 (1.943)
No high school	Reference cat.	Reference cat.	Reference cat.	Reference cat.
High school diploma	4.406 (5.066)	6.059 (5.723)	-3.759 (7.816)	-2.885 (7.711)
Some college	5.479 (4.965)	6.999 (5.607)	-3.086 (7.761)	-1.876 (7.659)
Bachelor's degree	7.512 (4.993)	8.726 (5.635)	-3.335 (7.777)	-2.502 (7.663)
Graduate degree	5.693 (5.054)	5.742 (5.687)	-3.517 (7.819)	-2.571 (7.711)
Income <$24.99K	Reference cat.	Reference cat.	Reference cat.	Reference cat.
Income $25–49.99K	1.800 (1.367)	0.753 (1.645)	1.879 (1.350)	1.605 (1.442)
Income $50–74.99K	2.779* (1.483)	1.604 (1.727)	3.568** (1.393)	2.942** (1.482)
Income $75–99.99K	8.392*** (1.879)	6.703*** (2.076)	6.540***(1.744)	6.188*** (1.827)
Income >$100K	8.945*** (1.835)	6.600*** (2.063)	8.323*** (1.691)	7.695*** (1.784)
Game/country dummies	Yes	Yes	Yes	Yes
Number of clusters	2,645	2,645	2,619	2,619
Number of obs.	10,580	10,580	13,095	13,095

Note: Presented are the unconditional marginal effects. The outcome variable is the winsorized *WTP* and includes 5,392 and 6,929 left-censored and 145 and 132 right-censored observations in the cup finals and league games sample, respectively. Standard errors clustered at the individual level are presented in parentheses. The IV-Tobit first stage is estimated with an ordered-probit regression, presented in Table A1 in Online Appendix A. ***, **, * denotes significance at the 1%, 5%, 10% levels, respectively.

Table 4. Predicted willingness-to-pay (WTP) for relocated soccer games.

	WTP for cup finals	WTP for league games
Average WTP	$19.95 ($24.01)	$17.99 ($19.43)
WTP for COO friendliness rating 1	$12.73 ($1.98)	$16.47 ($6.19)
WTP for COO friendliness rating 2	$18.62 ($6.31)	$15.06 ($7.70)
WTP for COO friendliness rating 3	$16.76 ($10.32)	$15.24 ($11.16)
WTP for COO friendliness rating 4	$20.27 ($21.93)	$18.45 ($19.05)
WTP for COO friendliness rating 5	$22.44 ($41.29)	$20.08 ($29.26)

Note: The WTP predictions for the cup finals and the league games are estimated by evaluating the expected effects from the Tobit models (and IV-Tobit models in parentheses) at the mean of the independent variables (from estimates in Table 3). *WTP* values for the country of origin (COO) friendliness ratings are estimated by setting the *country friendliness* variable to the respective rating level, holding all other variables at the mean.

Control variables

With regard to the control variables, we find that game-specific factors correlate with *WTP* for both types of games. As reported in Table 3, being a supporter of one of the two contestants or a supporter from another club of the same league is associated with a higher *WTP* for relocated games. Our measures of perceived game uncertainty point toward a U-shaped relation with *WTP*. This is in line with previous findings on European soccer (Buraimo & Simmons, 2015; Pawlowski et al., 2018; Nalbantis & Pawlowski, 2019), suggesting that spectators prefer games with a favorite instead of close games. Living in an area close to an MLS stadium does not seem to be related to *WTP*. With regard to socio-demographic characteristics, having a Hispanic background, being male, living in a larger household, and having comparably higher income are all positively associated with *WTP*. At the same time, however, *WTP* decreases with age. Finally, marital status, employment status, and educational attainment do not entail any statistically significant association with *WTP*.

Robustness checks

In order to test the sensitivity of our main findings, we conduct several robustness checks, considering alternative specifications of the estimation models and the *country friendliness* variable. The county distribution of *IV veteran share* shown in Figure B1 in Online Appendix B indicates that there is an overrepresentation of respondents in our samples living in low veteran share counties. In order to check whether our results are driven by these counties, we ran IV-Tobit subsample estimations (see column 1 and 3 in Table C1 in Online Appendix C) excluding respondents living in New York, Miami-Dade, and Queens, which are the three counties with the lowest veteran share (see Table B1 in Online Appendix B). Moreover, as the level of respondents' *country friendliness* may be correlated within each county, we re-estimate our IV-Tobit models with standard errors clustered at the county level (see column 2 and 4 in Table C1 in Online Appendix C). The results of both estimations remain similar compared to our main IV-models.

As the distribution of the Likert-scaled ordinal *country friendliness* variable is rather skewed (see Figure 2) and the true distance between the ordinal ratings is unknown, estimating differences in *WTP* between each rating level separately may inflate results. We thus follow previous studies (e.g. Masakure, 2016) and transform the ordinal measure

into a binary variable.[11] We first discriminate between individuals who perceive the respective country as unfriendly or neutral (ratings ≤ 3) and those who have friendly perception (ratings ≥ 4). In line with our main findings and independently of the used estimator we find a significantly higher *WTP* if the COO is perceived as friendly (ranging between $4 and $15, see Table C3 in Online Appendix C). When discriminating between individuals who perceive the respective country as unfriendly (ratings ≤ 2) and those who have a neutral or friendly perception (ratings ≥ 3), we still find a significant higher *WTP* if the COO is perceived as neutral or friendly. However, the overall effect becomes smaller (ranging between $2 and $10, see Table C4 in Online Appendix C) and the corresponding estimates are less precise. Likewise, the IV effects get smaller with relatively large standard errors. A similar pattern can be observed when excluding the neutral group from our estimations (see Table C5 in Online Appendix C).

Discussion and conclusion

Despite the increasing relevance of internationalization processes in sports in general and the potential impact of countries' political relations for spectator sports demand in particular, empirical research on this topic remains limited. This study is the first to test the role of perceived political relations between countries for the demand for relocated soccer games. More precisely, we focus on the "Big Five" European leagues as COO and the US as the importing country and use survey data to empirically test the relation between perceived friendliness toward the COO in international affairs and the demand for relocated soccer games.

Our findings suggest that perceived friendliness toward the COO is positively related to the WTP for relocated soccer games. We further account for potential endogeneity issues using an IV approach, which confirms our initial findings. Overall, we provide preliminary empirical evidence that the perceived political relations between importing country and COO are important for international demand of spectator sports.

These findings are in line with social identity theory (Tajfel & Turner, 1986) and its feature of in-group favoritism, which has been found to be particularly evident in agents'

[11] In case of the binary specifications, the CMP estimator estimates the first stage of the IV-Tobit models with a probit regression.

decision-making in international contexts (e.g. Krumer et al., 2022). Accordingly, sport consumers may use their national identity for in- and out-group formation if the COO becomes salient, affecting their individual preferences and demand (Akerlof & Kranton, 2000).

We also contribute to the rather limited literature on internationalization processes of professional sport leagues. As our study has shown, the impact of perceived political relations between countries on consumer demand may be particularly relevant for spectator sports – a product that is non-durable and easy to substitute – even in the absence of major political conflicts or crises. As such, the public perception about the COO in the importing country may constitute a crucial component for sport leagues' expansion strategies.

Our study also adds to the discussion on expansion strategies of professional sport leagues *in general* and European soccer leagues *in particular*. While the leagues' target market selection seems to be mainly driven by macro segmentation (e.g. market size and potential), micro segmentation (i.e. based on consumers' personal values) is also a key criterion in international market selection (Gaston-Breton & Martín, 2011). In this line, our findings suggest that sport consumers' political opinions about the COO should be considered in the leagues' expansion strategies to select and approach new markets. Such information is readily available for sport policy makers and managers as country ratings are frequently surveyed in opinion polls (e.g. Gallup Poll). Depending on the segmentation approach, decision-makers may consider differences in political opinions of certain target groups and geographic markets (e.g. regional or national markets) to maximize revenues. For instance, choosing a stadium located in a municipality where the residents perceive the COO as friendly seems advisable because expected ticketing revenues are larger. Moreover, we expect that any indirect effect, i.e. stimulating demand for games of the corresponding clubs and leagues or attracting new "satellite-fans" in the longer run, potentially works better when the COO is perceived as friendly. Testing the latter, however, is beyond the scope of this paper.

Although our study adds new insights on the international demand for spectator sports, it also has some limitations. *First*, we cannot fully rule out the possibility that the exclusion restriction is violated in our IV approach. Although this is not testable, more information

on the veteran status of our survey respondents or potential spillover effects would increase the validity of our IV approach. Thus, this paper provides a first step to explore the role of political relations for sports demand but is not able to provide conclusive causal evidence. *Second*, we only examine the demand for soccer in the US, but it also seems important to look at potential cross-country differences as the national identity varies across countries, shaping individual preferences (e.g. Mayda & Rodrik, 2005). *Third*, although our scenario approach has some advantages to identify the relation of interest, its hypothetical nature also includes some drawbacks (e.g. selection of games, WTP elicitation method). As the number of relocated games continues to increase, future research could take advantage of larger samples of stadium attendance data and combine them with data from opinion polls, measuring countries' perceived political relations at the national or regional level. *Finally*, although we focus on the impact of political relations on demand for relocated games, it would also be plausible to assume that relocated games may affect consumer friendliness toward the COO (e.g. Abdi et al., 2018). Thus, it seems promising to empirically examine whether sport events can serve as a policy instrument for diplomatic purposes.

References

Abdi, K., Talebpour, M., Fullerton, J., Ranjkesh, M. J., & Jabbari Nooghabi, H. (2018). Converting sports diplomacy to diplomatic outcomes: Introducing a sports diplomacy model. *International Area Studies Review*, *21*(4), 365–381. https://doi.org/10.1177/2233865918808058

Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, *70*(2), 542–555. https://doi.org/10.1017/S0022381608080493

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753. https://doi.org/10.1162/003355300554881

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010

Becker, S. O. (2016). Using instrumental variables to establish causality. *IZA World of Labor*, (250), 1–10. https://doi.org/10.15185/izawol.250

Bennett, R., Ali-Choudhury, R., & Mousley, W. (2007). Television viewers' motivations to follow the 2005 Ashes Test series: Implications for the rebranding of English cricket. *Journal of Product & Brand Management*, *16*(1), 23–37. https://doi.org/10.1108/10610420710731133

Buraimo, B., & Simmons, R. (2015). Uncertainty of outcome or star quality? Television audience demand for English Premier League football. *International Journal of the Economics of Business*, *22*(3), 449–469. https://doi.org/10.1080/13571516.2015.1010282

Cáceres-Delpiano, J., De Moragas, A. I., Facchini, G., & González, I. (2021). Intergroup contact and nation building: Evidence from military service in Spain. *Journal of Public Economics*, *201*, https://doi.org/10.1016/j.jpubeco.2021.104477. https://doi.org/10.1016/j.jpubeco.2021.104477

Castellanos-Garcia, P., García, J., & Sánchez , J. M. (2011). The willingness to pay to keep a football club in a city: how important are the methodological issues? *Journal of Sports Economics*, *12(*4), 464–486. https://doi.org/10.1177/1527002510385301

Chavis, L., & Leslie, P. (2009). Consumer boycotts: The impact of the Iraq war on French wine sales in the US. *Quantitative Marketing and Economics*, *7*(1), 37–67. https://doi.org/10.1007/s11129-008-9043-y

Funk, C. A., Arthurs, J. D., Treviño, L. J., & Joireman, J. (2010). Consumer animosity in the global value chain: The effect of international production shifts on willingness to purchase hybrid products. *Journal of International Business Studies*, *41*(4), 639–651. https://doi.org/10.1057/jibs.2009.29

Gaston-Breton, C., & Martín, O. M. (2011). International market selection and segmentation: A two-stage model. *International Marketing Review*, *28*(3), 267–290. https://doi.org/10.1108/02651331111132857

Ghosh, D., & Vogt, A. (2012). Outliers: An evaluation of methodologies. *Joint Statistical Meetings*, 3455–3460. http://www.asasrms.org/Proceedings/y2012/Files/304068_72402.pdf

Goetz, S. J., Davlasheridze, M., Han, Y., & Fleming-Muñoz, D. A. (2019). Explaining the 2016 vote for president Trump across us counties. *Applied Economic Perspectives and Policy*, *41*(4), 703–722. https://doi.org/10.1093/aepp/ppy026

Green, D., Jacowitz, K. E., Kahneman, D., & McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, *20*(2), 85–116. https://doi.org/10.1016/S0928-7655(97)00031-6

Heilmann, K. (2016). Does political conflict hurt trade? Evidence from consumer boycotts. *Journal of International Economics*, *99*, 179–191. https://doi.org/10.1016/j.jinteco.2015.11.008

Johnson, B. K., Groothuis, P. A., & Whitehead, J. C. (2001). The value of public goods generated by a major league sports team: The CVM approach. *Journal of Sports Economics*, *2*(1), 6–21. https://doi.org/10.1177/152700250100200102

Johnson, B. K., Mondello, M. J., & Whitehead, J. C. (2007). The value of public goods generated by a National Football League team. *Journal of Sport Management*, *21*(1), 123–136. https://doi.org/10.1123/jsm.21.1.123

Johnson, B. K., & Whitehead, J. C. (2000). Value of public goods from sports stadiums: The CVM approach. *Contemporary Economic Policy*, *18*(1), 48–58. https://doi.org/10.1111/j.1465-7287.2000.tb00005.x

Kim, S., Greenwell, T. C., Andrew, D. P., Lee, J., & Mahony, D. F. (2008). An analysis of spectator motives in an individual combat sport: A study of mixed martial arts fans. *Sport Marketing Quarterly*, *17*(2), 109–119. https://oaks.kent.edu/flapubs/17

Klein, J. G., Ettenson, R., & Morris, M. D. (1998). The animosity model of foreign product purchase: An empirical test in the People's Republic of China. *Journal of Marketing*, *62*(1), 89–100. https://doi.org/10.1177/002224299806200108

Konjer, M., Meier, H. E., & Wedeking, K. (2017). Consumer demand for telecasts of tennis matches in Germany. *Journal of Sports Economics*, *18*(4), 351–375. https://doi.org/10.1177/1527002515577882

Krueger, J. S., & Pedraza, F. I. (2012). Missing voices: War attitudes among military service-connected civilians. *Armed Forces & Society*, *38*(3), 391–412. https://doi.org/10.1177/0095327X11428786

Krumer, A., Otto, F., & Pawlowski, T. (2022) Nationalistic bias among international experts: Evidence from professional ski jumping. *The Scandinavian Journal of Economics*, *124*(1), 278–300. https://doi.org/10.1111/sjoe.12451

Leong, S. M., Cote, J. A., Ang, S. H., Tan, S. J., Jung, K., Kau, A. K., & Pornpitakpan, C. (2008). Understanding consumer animosity in an international crisis: nature, antecedents, and consequences. *Journal of International Business Studies*, *39*(6), 996–1009. https://doi.org/10.1057/palgrave.jibs.8400392

Lo, A. Y., & Jim, C. Y. (2015). Protest response and willingness to pay for culturally significant urban trees: Implications for Contingent Valuation Method.

Ecological Economics, *114*, 58–66. https://doi.org/10.1016/j.ecolecon.2015.03.012

Loschky, J., & Riffkin, R. (2015). *Canada, Great Britain are Americans' most favored nations*. https://news.gallup.com/poll/181961/canada-great-britain-americans-favored-nations.aspx.

Masakure, O. (2016). The effect of employee loyalty on wages. *Journal of Economic Psychology*, *56*, 274–298. https://doi.org/10.1016/j.joep.2016.08.003

Mayda, A. M., & Rodrik, D. (2005). Why are some people (and countries) more protectionist than others? *European Economic Review*, *49*(6), 1393–1430. https://doi.org/10.1016/j.euroecorev.2004.01.002

McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. *Review of Economics and Statistics*, *62*(2), 318–321. https://doi.org/10.2307/1924766

Miller, K. M., Hofstetter, R., Krohmer, H., & Zhang, Z. J. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, *48*(1), 172–184. https://doi.org/10.1509/jmkr.48.1.172

Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, *30*(3), 313–325. https://doi.org/10.1007/s10640-004-3332-z

Nalbantis, G., & Pawlowski, T. (2016). *The Demand for International Football Telecasts in the United States*. Palgrave Macmillan.

Nalbantis, G., & Pawlowski, T. (2019). US demand for European soccer telecasts: a between-country test of the uncertainty of outcome hypothesis. *Journal of Sports Economics*, *20*(6), 797–818. https://doi.org/10.1177/1527002518817598

Nalbantis, G., Pawlowski, T., & Coates, D. (2017). The fans' perception of competitive balance and its impact on willingness-to-pay for a single game. *Journal of Sports Economics*, *18*(5), 479–505. https://doi.org/10.1177/1527002515588137

Nalbantis, G., Pawlowski, T., & Schreyer, D. (2021). Substitutions effects and the transnational demand for European soccer telecasts (Working paper). University of Tübingen.

Nüesch, S., & Franck, E. (2009). The role of patriotism in explaining the TV audience of national team games – evidence from four international tournaments. *Journal of Media Economics*, *22*(1), 6–19. https://doi.org/10.1080/08997760902724472

Pandya, S. S., & Venkatesan, R. (2016). French roast: consumer response to international conflict – Evidence from supermarket scanner data. *Review of Economics and Statistics*, *98*(1), 42–56. https://doi.org/10.1162/REST_a_00526

Parameswaran, R., & Yaprak, A. (1987). A cross-national comparison of consumer research measures. *Journal of International Business Studies*, *18*(1), 35-49. https://doi.org/10.1057/palgrave.jibs.8490398

Pawlowski, T., Nalbantis, G., & Coates, D. (2018). Perceived game uncertainty, suspense and the demand for sport. *Economic Inquiry*, *56*(1), 173–192. https://doi.org/10.1111/ecin.12462

Rand, D. G., Pfeiffer, T., Dreber, A., Sheketoff, R. W., Wernerfelt, N. C., & Benkler, Y. (2009). Dynamic remodeling of in-group bias during the 2008 presidential election. *Proceedings of the National Academy of Sciences*, *106*(15), 6187–6191. https://doi.org/10.1073/pnas.0811552106

Riefler, P., & Diamantopoulos, A. (2007). Consumer animosity: a literature review and a reconsideration of its measurement. *International Marketing Review*, *24*(1), 87–119. https://doi.org/10.1108/02651330710727204

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal*, *11*(2), 159–206. https://doi.org/10.1177/1536867X1101100202

Sande, J. B., & Ghosh, M. (2018). Endogeneity in survey research. *International Journal of Research in Marketing*, *35*(2), 185–204. https://doi.org/10.1016/j.ijresmar.2018.01.005

Schläpfer, F., & Fischhoff, B. (2012). Task familiarity and contextual cues predict hypothetical bias in a meta-analysis of stated preference studies. *Ecological Economics*, *81*, 44–47. https://doi.org/10.1016/j.ecolecon.2012.06.016

Schreyer, D., Schmidt, S. L., & Torgler, B. (2018). Game outcome uncertainty in the English Premier League: Do German fans care? *Journal of Sports Economics*, *19*(5), 625–644. https://doi.org/10.1177/1527002516673406

Shoham, A., Davidow, M., Klein, J. G., & Ruvio, A. (2006). Animosity on the home front: The Intifada in Israel and its impact on consumer behavior. *Journal of International Marketing*, *14*(3), 92–114. https://doi.org/10.1509/jimk.14.3.92

Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature*, *41*(4), 1137–1187. https://doi.org/10.1257/jel.41.4.1137

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7–24). Nelson-Hall.

Teachman, J. (2013). A note on disappearing veterans. *Armed Forces & Society*, *39*(4), 740–750. https://doi.org/10.1177/0095327X12468731

United States Census Bureau (2020). *2011-2015 American Community Survey 5-Year Estimates*. https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2015/5-year.html

Usunier, J. C., & Cestre, G. (2007). Product ethnicity: Revisiting the match between products and countries. *Journal of International Marketing*, *15*(3), 32–72. https://doi.org/10.1509/jimk.15.3.32

Van Reeth, D. (2013). Television demand for the Tour de France: The importance of outcome uncertainty, patriotism and doping. *International Journal of Sport Finance*, *8*, 39–60.

Appendix (online appendix)

Appendix A: First-stage results of IV-Tobit estimations

Table A1. First-stage results of IV-Tobit estimations (second-stage results in Table 3).

	Cup finals	League games
IV veteran share (ACS)	-2.209*** (0.573)	-1.355** (0.605)
Club supporter (game)	0.238*** (0.075)	0.281*** (0.061)
Club supporter (league)	0.347*** (0.045)	0.382*** (0.045)
Perceived team A win Pr	-0.175*** (0.023)	-0.212*** (0.022)
Perceived team A win Pr²	0.015*** (0.002)	0.018*** (0.002)
Nearby MLS stadium	-0.044 (0.059)	-0.034 (0.064)
Hispanophone	0.090 (0.062)	0.115** (0.058)
White	0.021 (0.049)	0.059 (0.051)
Male	0.075** (0.035)	0.086** (0.035)
Married	-0.008 (0.041)	-0.063 (0.042)
Household size	-0.005 (0.016)	0.022 (0.017)
Age (in years)	0.004*** (0.001)	0.005*** (0.001)
Unemployed	-0.010 (0.082)	0.066 (0.084)
No high school	Reference cat.	Reference cat.
High school diploma	-0.153 (0.267)	-0.153 (0.284)
Some college	-0.132 (0.264)	-0.213 (0.280)
Bachelor's degree	-0.101 (0.264)	-0.142 (0.280)
Graduate degree	0.023 (0.266)	-0.161 (0.282)
Income <$24.99K	Reference cat.	Reference cat.
Income $25–49.99K	0.129** (0.063)	0.082 (0.065)
Income $50–74.99K	0.137** (0.064)	0.155** (0.066)
Income $75–99.99K	0.187*** (0.072)	0.102 (0.074)
Income >$100K	0.246*** (0.071)	0.148** (0.072)
Cut point 1 (μ1)	-2.355 (0.294)	-2.478 (0.307)
Cut point 2 (μ2)	-1.720 (0.291)	-1.813 (0.302)
Cut point 3 (μ3)	-0.638 (0.289)	-0.689 (0.301)
Cut point 4 (μ4)	0.360 (0.289)	0.352 (0.301)
Game/country dummies	Yes	Yes
Number of clusters	2,645	2,619
Number of obs.	10,580	13,095

Note: Presented are the estimated coefficients of the first-stage ordered-probit regressions of the IV-Tobit estimations, presented in Table 3. The endogenous variable is the ordinally scaled country friendliness ranging from 1-5. Standard errors clustered at the individual level are presented in parentheses. ***, **, * denotes significance at the 1%, 5%, 10% level, respectively.

Appendix B: Distribution of IV veteran share per county

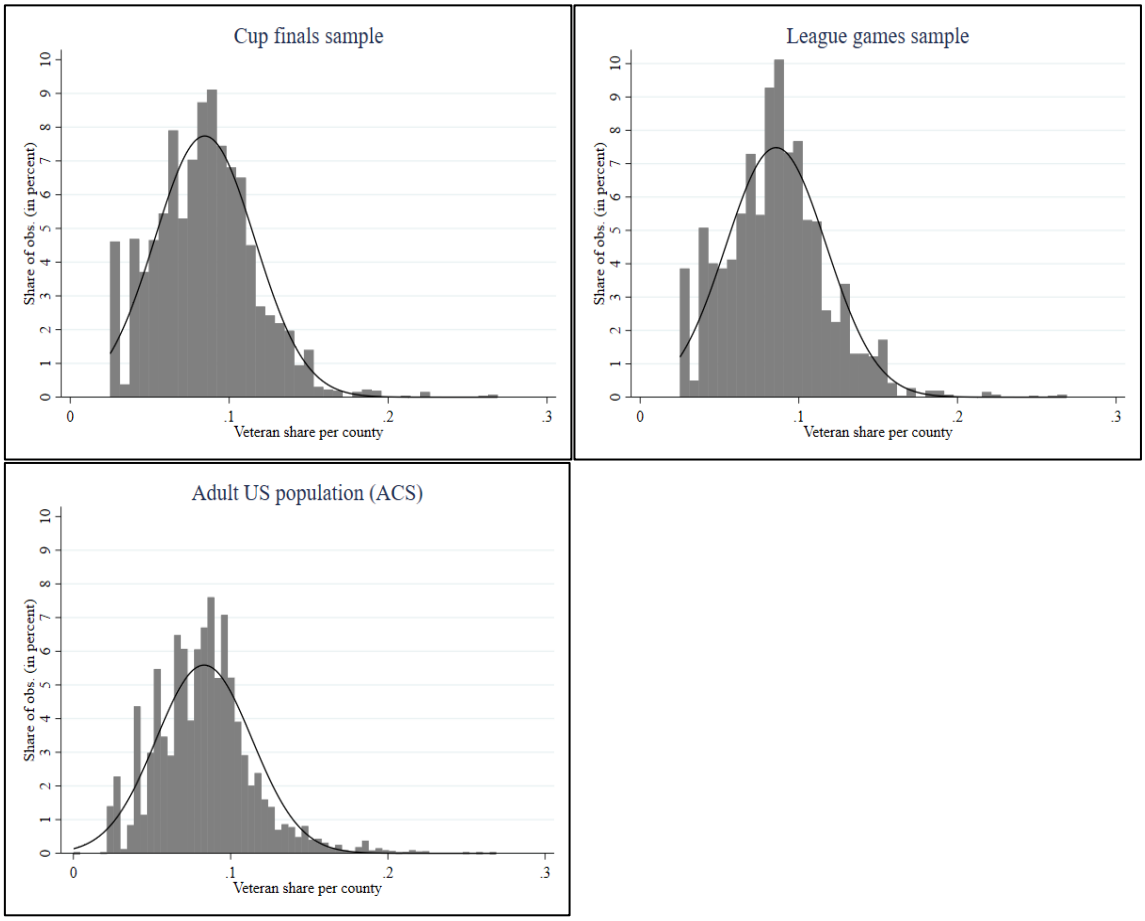

Figure B1. Distribution of the veteran share per county.

Note: Adult US population data stems from American Community Survey (ACS).

Table B1. Counties with lowest and highest veteran share per county.

Bottom 10 counties	Veteran share	*N*	Top 10 counties	Veteran share	*N*
Cup finals sample					
New York County, New York	0.025	70	Virginia Beach city, Virginia	0.186	5
Miami-Dade County, Florida	0.027	25	Comanche County, Oklahoma	0.189	1
Queens County, New York	0.027	17	Cochise County, Arizona	0.193	3
Hudson County, New Jersey	0.028	2	Stafford County, Virginia	0.193	1
Bronx County, New York	0.029	7	Coryell County, Texas	0.194	1
Maverick County, Texas	0.030	1	Island County, Washington	0.212	1
San Francisco County, Cali.	0.036	9	Onslow County, North Caroli.	0.222	2
Holmes County, Mississippi	0.037	1	Okaloosa County, Florida	0.223	2
Grant County, Kansas	0.039	1	Pulaski County, Missouri	0.257	1
Passaic County, New Jersey	0.039	4	Elmore County, Idaho	0.269	2
League games sample					
New York County, New York	0.025	58	Virginia Beach city, Virginia	0.186	5
Miami-Dade County, Florida	0.027	25	Sumter County, Florida	0.190	1
Queens County, New York	0.027	13	Cochise County, Arizona	0.193	3
Hudson County, New Jersey	0.028	3	Stafford County, Virginia	0.193	1
Bronx County, New York	0.029	2	Bell County, Texas	0.215	1
Harrisonburg city, Virginia	0.032	1	Onslow County, North Carolina	0.222	1
Webb County, Texas	0.033	1	Okaloosa County, Florida	0.223	2
San Francisco County, Cali.	0.036	11	Liberty County, Georgia	0.248	2
Holmes County, Mississippi	0.037	1	Pulaski County, Missouri	0.257	1
Moore County, Texas	0.038	3	Elmore County, Idaho	0.269	2

Appendix C: Robustness checks

Table C1. IV-Tobit estimates of country friendliness.

	Cup finals		League games	
	(1)	(2)	(3)	(4)
Country friendliness 1	-7.961*** (0.727)	-8.334*** (0.833)	-4.270** (2.053)	-4.971*** (1.843).
Country friendliness 2	-3.945*** (0.717)	-4.005*** (0.717)	-3.160*** (1.010)	-3.465*** (0.957)
Country friendliness 3	Reference cat.	Reference cat.	Reference cat.	Reference cat.
Country friendliness 4	11.322*** (1.359)	11.616*** (1.744)	7.708*** (1.609)	7.884*** (1.551)
Country friendliness 5	31.117*** (5.000)	30.973*** (6.252)	16.946*** (5.100)	18.093*** (4.887)
Club supporter (game)	12.967*** (2.665)	14.718*** (2.720)	17.634*** (2.229)	15.359*** (2.348)
Club supporter (league)	9.783*** (1.290)	10.475*** (1.363)	13.123*** (1.588)	13.359*** (1.462)
Perceived team A win Pr	-3.884*** (0.612)	-4.086*** (0.605)	-3.201*** (0.589)	-3.325*** (0.529)
Perceived team A win Pr²	0.359*** (0.058)	0.378*** (0.057)	0.285*** (0.056)	0.296*** (0.052)
Nearby MLS stadium	-3.425* (1.795)	-1.028 (2.456)	0.089 (1.603)	1.420 (1.795)
Hispanophone	2.908* (1.749)	3.250* (1.784)	5.525*** (1.827)	5.153*** (1.495)
White	-6.114*** (1.356)	-5.900*** (1.399)	-2.182* (1.239)	-2.124* (1.236)
Male	1.764* (1.021)	1.779** (0.880)	1.771* (0.914)	2.032** (0.906)
Married	-1.049 (1.200)	-0.783 (1.138)	-0.312 (1.087)	-0.113 (1.076)
Household size	0.951** (0.447)	0.734 (0.476)	1.445*** (0.482)	1.465*** (0.435)
Age (in years)	-0.433*** (0.042)	-0.455*** (0.401)	-0.349*** (0.040)	-0.365*** (0.042)
Unemployed	-2.253 (1.931)	-2.655 (2.047)	-1.084 (1.894)	-1.154 (2.064)
No high school	Reference cat.	Reference cat.	Reference cat.	Reference cat.
High school diploma	5.919 (5.954)	6.059 (5.529)	-3.220 (7.685)	-2.885 (7.910)
Some college	7.749 (5.864)	6.999 (5.585)	1.848 (7.641)	-1.876 (7.678)
Bachelor's degree	9.647 (5.897)	8.726 (5.706)	-2.748 (7.642)	-2.502 (7.632)
Graduate degree	6.339 (5.948)	5.742 (5.735)	-2.653 (7.694)	-2.571 (7.814)
Income <$24.99K	Reference cat.	Reference cat.	Reference cat.	Reference cat.
Income $25–49.99K	0.504 (1.631)	0.753 (1.545)	1.695 (1.407)	1.605 (1.306)
Income $50–74.99K	1.383 (1.708)	1.604 (1.644)	2.995** (1.445)	2.942** (1.459)
Income $75–99.99K	6.616*** (2.075)	6.703*** (2.378)	6.419*** (1.809)	6.188*** (1.813)
Income >$100K	6.051*** (2.035)	6.601*** (1.989)	7.572*** (1.766)	7.695*** (1.699)
Game/country dummies	Yes	Yes	Yes	Yes
Number of clusters	2,533	680	2,523	673
Number of obs.	10,132	10,580	12,615	13,095

Note: Presented are the unconditional marginal effects. The outcome variable is the winsorized *WTP*. Columns 1 and 3 present subsample estimations, excluding observations from the three counties with the lowest veteran share, and including 5,323 and 6,847 left-censored and 116 and 115 right-censored *WTP* observations in the cup finals and league games sample, respectively. Standard errors are presented in parentheses and clustered at the individual level. In columns 2 and 4, *WTP* includes 5,392 and 6,929 left-censored and 145 and 132 right-censored observations in the cup finals and league games sample, respectively. Here, standard errors are clustered at the county level. The IV-Tobit first stage is estimated with an ordered-probit regression, presented in Table C2. ***, **, * denotes significance at the 1%, 5%, 10% levels, respectively.

Table C2. First-stage results of IV-Tobit estimations (second-stage results in Table C1).

	Cup finals		League games	
	(1)	(2)	(3)	(4)
IV veteran share (ACS)	-2.487*** (0.588)	-2.209*** (0.720)	-1.340** (0.633)	-1.355** (0.567)
Club supporter (game)	0.285*** (0.081)	0.238*** (0.089)	0.308*** (0.066)	0.281*** (0.058)
Club supporter (league)	0.335*** (0.047)	0.347*** (0.043)	0.393*** (0.046)	0.382*** (0.047)
Perceived team A win Pr	-0.172*** (0.023)	-0.175*** (0.022)	-0.211*** (0.023)	-0.212*** (0.022)
Perceived team A win Pr²	0.015*** (0.002)	0.015*** (0.002)	0.018*** (0.002)	0.018*** (0.002)
Nearby MLS stadium	-0.042 (0.066)	-0.044 (0.055)	-0.029 (0.074)	-0.034 (0.055)
Hispanophone	0.104 (0.066)	0.090 (0.063)	0.106* (0.062)	0.115** (0.058)
White	0.007 (0.050)	0.021 (0.054)	0.053 (0.052)	0.059 (0.048)
Male	0.075** (0.036)	0.075** (0.034)	0.083** (0.035)	0.086*** (0.033)
Married	-0.018 (0.042)	-0.008 (0.040)	-0.062 (0.043)	-0.063 (0.043)
Household size	-0.008 (0.016)	-0.005 (0.015)	0.020 (0.017)	0.022 (0.015)
Age (in years)	0.004** (0.002)	0.004*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
Unemployed	-0.026 (0.082)	-0.010 (0.086)	0.069 (0.084)	0.066 (0.083)
No high school	Reference cat.	Reference cat.	Reference cat.	Reference cat.
High school diploma	-0.137 (0.288)	-0.153 (0.265)	-0.142 (0.284)	-0.153 (0.282)
Some college	-0.126 (0.285)	-0.132 (0.264)	-0.217 (0.280)	-0.213 (0.279)
Bachelor's degree	-0.102 (0.286)	-0.101 (0.265)	-0.132 (0.281)	-0.142 (0.282)
Graduate degree	-0.026 (0.288)	0.023 (0.267)	-0.173 (0.282)	-0.161 (0.285)
Income <$24.99K	Reference cat.	Reference cat.	Reference cat.	Reference cat.
Income $25–49.99K	0.122* (0.063)	0.129* (0.067)	0.078 (0.066)	0.082 (0.062)
Income $50–74.99K	0.123* (0.065)	0.137** (0.068)	0.148** (0.066)	0.155** (0.063)
Income $75–99.99K	0.182** (0.073)	0.187*** (0.079)	0.106 (0.074)	0.102 (0.071)
Income >$100K	0.247 *** (0.072)	0.246*** (0.074)	0.154** (0.073)	0.148** (0.069)
Cut point 1 (μ1)	-2.423 (0.314)	-2.355 (0.293)	-2.487 (0.308)	-2.478 (0.295)
Cut point 2 (μ2)	-1.784 (0.311)	-1.720 (0.289)	-1.821 (0.303)	-1.813 (0.288)
Cut point 3 (μ3)	-0.700 (0.309)	-0.638 (0.290)	-0.694 (0.302)	-0.689 (0.290)
Cut point 4 (μ4)	0.300 (0.310)	0.360 (0.290)	0.341 (0.302)	0.352 (0.289)
Game/country dummies	Yes	Yes	Yes	Yes
Number of clusters	2,533	680	2,523	673
Number of obs.	10,132	10,580	12,615	13,095

Note: Presented are the estimated coefficients of the first-stage ordered-probit regressions of the IV-Tobit estimations, presented in Table C1. The endogenous variable is the ordinally scaled *country friendliness* ranging from 1-5. Columns 1 and 3 present subsample estimations, excluding observations from the three counties with the lowest veteran share. Standard errors are presented in parentheses and are clustered at the individual level. In columns 2 and 4, standard errors are clustered at the county level. ***, **, * denotes significance at the 1%, 5%, 10% level, respectively.

Table C3. Estimates of country friendliness dummy specification: ratings 1-3 vs. 4-5.

	Cup finals		League games	
Estimation model	Instrument effect	CF estimate	Instrument effect	CF estimate
Tobit model		4.395*** (0.774)		3.827*** (0.714)
IV-Tobit model	-0.785*** (0.242)	15.017*** (1.974)	-0.562** (0.230)	12.396*** (2.255)
S.E. clustered at county level	-0.785** (0.303)	15.017*** (2.658)	-0.562** (0.247)	12.396*** (2.242)
No. of obs. (left-/right censored obs.)	10,580 (5,392/145)		13,095 (6,929/132)	
Excl. low veteran share counties	-0.897*** (0.253)	14.517*** (2.049)	-0.535** (0.240)	11.063*** (2.428)
No. of obs. (left-/right censored obs.)	10,132 (5,323/116)		12,615 (6,847/115)	

Note: Each line represents a regression model, including all control variables. Presented are the average marginal effects for the instrument *IV veteran share* in a first-stage probit model and the unconditional marginal effects of *country friendliness* (CF) in the Tobit and IV-Tobit models (second stage). The outcome variable is the winsorized *WTP*. Standard errors (S.E.) (clustered at individual level if not stated otherwise) are presented in parentheses. ***, **, * denotes significance at the 1%, 5%, 10% levels, respectively.

Table C4. Estimates of country friendliness dummy specification: ratings 1-2 vs. 3-5.

	Cup finals		League games	
Estimation model	Instrument effect	CF estimate	Instrument effect	CF estimate
Tobit model		2.820** (1.224)		2.477** (1.187)
IV-Tobit model	-0.142 (0.140)	9.457*** (2.036)	-0.160 (0.118)	7.485*** (2.037)
S.E. clustered at county level	-0.142 (0.145)	9.457*** (1.768)	-0.160 (0.115)	7.485*** (1.973)
No. of obs. (left-/right censored obs.)	10,580 (5,392/145)		13,095 (6,929/132)	
Excl. low veteran share counties	-0.197 (0.144)	9.871*** (1.766)	-0.132 (0.126)	6.452*** (2.176)
No. of obs. (left-/right censored obs.)	10,132 (5,323/116)		12,615 (6,847/115)	

Note: Each line represents a regression model, including all control variables. Presented are the average marginal effects for the instrument *IV veteran share* in a first-stage probit model and the unconditional marginal effects of *country friendliness* (CF) in the Tobit and IV-Tobit models (second stage). The outcome variable is the winsorized *WTP*. Standard errors (S.E.) (clustered at individual level if not stated otherwise) are presented in parentheses. ***, **, * denotes significance at the 1%, 5%, 10% levels, respectively.

Table C5. Estimates of country friendliness dummy specification ratings: 1-2 vs. 4-5 (subsample without rating 3).

	Cup finals		League games	
Estimation model	Instrument effect	CF estimate	Instrument effect	CF estimate
Tobit model		4.404*** (1.342)		4.093*** (1.262)
IV-Tobit model	-0.257 (0.184)	11.989*** (2.058)	-0.286* (0.158)	9.915*** (2.185)
S.E. clustered at county level	-0.257 (0.196)	11.989*** (1.823)	-0.286* (0.146)	9.915*** (2.180)
No. of obs. (left-/right censored obs.)	7,803 (3,683/ 128)		9,547 (4,669/100)	
Excl. low veteran share counties	-0.336* (0.190)	11.980*** (1.873)	-0.253 (0.168)	8.687*** (2.253)
No. of obs. (left-/right censored obs.)	7,449 (3,643/102)		9,156 (4,607/84)	

Note: Each line represents a regression model, including all control variables. Presented are the average marginal effects for the instrument *IV veteran share* in a first-stage probit model and the unconditional marginal effects of *country friendliness* (CF) in the Tobit and IV-Tobit models (second stage). The outcome variable is the winsorized *WTP*. Standard errors (S.E.) (clustered at individual level if not stated otherwise) are presented in parentheses. ***, **, * denotes significance at the 1%, 5%, 10% levels, respectively.

4.3 Nationalistic bias among international experts: Evidence from professional ski jumping*

Scand. J. of Economics 124(1), 278–300, 2022
DOI: 10.1111/sjoe.12451

Nationalistic bias among international experts: evidence from professional ski jumping*

Alex Krumer

Molde University College, NO-6402 Molde, Norway
alex.krumer@himolde.no

Felix Otto

University of Tübingen, DE-72074 Tübingen, Germany
felix.otto@uni-tuebingen.de

Tim Pawlowski

University of Tübingen, DE-72074 Tübingen, Germany
tim.pawlowski@uni-tuebingen.de

Abstract

Ski jumping competitions involve subjective evaluations by judges from different countries. This might lead to nationalistic bias, according to which judges assign higher scores to their compatriots. To test this claim empirically, we exploit within-performance variation of scores from all World Cup, World Championship, and Olympic Games competitions between the 2010/11 and 2016/17 seasons. Our findings confirm that judges assign significantly higher scores to their compatriots. The magnitude of this nationalistic bias is significantly higher in more corrupt countries. We do not find that judges assign significantly different scores to jumpers whose compatriots are present on the judging panel.

Keywords: In-group favoritism; judging panel; nationalistic bias; replication study; subjective performance evaluation

JEL classification: *D*71; *D*91; *Z*20

*Acknowledgements: We thank international ski jumping judge, Ole Walseth, the Olympic medalist and former world record holder, Johan Remen Evensen, as well as Kjetil K. Haugen and Geir Oterhals, for providing valuable background knowledge on professional ski jumping.

Nationalistic bias among international experts: Evidence from professional ski jumping

Introduction

Can well-trained and professional experts resist inherent preferences toward in-group members in their subjective evaluations? Do these experts use strategic motives when they evaluate in-group members of their counterparts? We try to answer these questions by studying the subjective evaluations of a panel of international experts who evaluate the performance of highly skilled professionals in real-life tournament settings with high monetary rewards.

In general, in-group favoritism based on the division of people into groups, according to some predefined rule, is a very well-established phenomenon. For example, Efferson et al. (2008) showed that even different signs on shirts were enough to divide people into groups and create in-group favoritism, according to which members of one group favor in-group members over out-group members. Thus, it is likely that in-group favoritism is one of the more primitive human instincts that developed during the evolutionary process (Sumner, 1907; Yuki, 2003), whose effects can even be observed in neurological processes in our brain.[1]

In-group favoritism has also been documented in various non-experimental settings. For example, Shayo and Zussman (2011) found that legal claims are more likely to be accepted if the judge and the plaintiff have the same ethnicity. Spierdijk and Vellekoop (2009) showed in-group favoritism based on geographical proximity in Eurovision Song Contests. Several other studies have shown evidence of favoritism in professional sport. For example, Price and Wolfers (2010) found that NBA players have fewer fouls called against them when their race matches that of the refereeing crew. Similarly, Pope and Pope (2015) demonstrated that referees favor their compatriot players by assigning them more beneficial foul calls in UEFA Champions League games. Very recently, Faltings et

[1] As evidence, Andrews et al. (2019) tested the brain activities of fans from two rival teams who watched the same soccer game. They found a correlation between supporters of the same team in brain activities in areas that are known to be active in reward, self-identity, and control of movement. However, these brain activities were significantly different between the two groups of fans.

al. (2021) investigated Swiss soccer matches and showed that referees from one linguistic group assign significantly more yellow and red cards to teams from a different linguistic area.

In this paper, we build on the efforts of Zitzewitz (2006), who studied the subjective evaluation by judges in professional ski jumping based on data from 25 competitions in 2002. In these competitions, jumpers maximize their aggregate point score, which is determined by jumping distance (an objectively measured performance) and style points (a subjectively measured performance). His findings were striking: using within-performance (jump) variation of scores, he showed that judges assigned a significantly larger number of style points to their compatriot jumpers than the other judges who observed the same performance. In addition, Zitzewitz (2006) found a similar pattern of a nationalistic bias in figure skating competitions. Using a similar estimation strategy, Sandberg (2018) showed an analogous result in dressage competitions. More recently, Scholten et al. (2020) used data from 41 World Cup ski jumping competitions, and provided suggestive evidence that nationalistic bias is still present. However, the number of events analyzed in their study is limited and the estimation strategy employed misses several key issues, making any comparison with the early findings by Zitzewitz (2006) impossible.[2]

We replicate and extend the analyses by Zitzewitz (2006) using data on 76,775 different evaluations of ski jumping judges from 203 competitions, comprising all the World Cups, Nordic World Ski Championships, and the Olympic Games between 2010/11 and 2016/17. Such an exercise seems highly relevant for two reasons. First, there is a growing consensus about the importance of replication studies in science.[3] Second, it seems highly

[2] Scholten et al. (2020) neither exploited the within-performance (jump) variation of scores nor controlled for competition fixed effects. Moreover, they did not investigate the possibility of a compensating bias, according to which judges assign different scores to jumpers whose compatriots are present on the judging panel.

[3] For example, Open Science Collaboration (2015) replicated the results of only 36 out of 100 experimental and correlational studies that were published in top academic journals in psychology. In the same vein, Silberzahn et al. (2018) showed a high variance in the results of 29 scientific teams that investigated the same dataset, highlighting the importance of crowdsourced research that "can balance discussions, validate scientific findings and better inform policymakers" (Silberzahn and Uhlmann, 2015, p. 190). Finally, Ioannidis and Doucouliagos (2013) discussed the empirical evidence on the lack of a robust reproducibility culture in economics and business research. Therefore, replication of original findings is an important scientific task.

relevant from a policy perspective to see whether problems that had been identified before have been solved over time.[4]

Comparing the score of a compatriot judge to scores of the other members of the panel within each jump, we find that compatriot judges assign close to 0.09 points more compared to their counterparts. This is equivalent to 29 percent of the within-jump standard deviation of scores, a non-negligible difference. As such, the nationalistic bias in professional ski jumping is remarkably persistent and still exists more than a decade after the initial findings by Zitzewitz (2006), which were also featured in the media.[5]

Further analysis suggests that the nationalistic bias is higher in more corrupt countries. Out of the 12 most observed countries in our data, only Norway and Finland had negligible estimates of a nationalistic bias, both statistically and economically. In contrast, Russian judges assigned, on average, 0.22 points more to Russian jumpers than the other judges on the panel.[6]

Finally, we test whether there is evidence of strategic voting, according to which judges assign different scores to jumpers whose compatriots are present on the judging panel. The evidence of such a strategic voting is mixed. On the one hand, Zitzewitz (2006), who coined the term "compensating bias" for that phenomenon, found that, for some specifications, the ski jumping judges assign significantly lower scores to jumpers if the other judge on the panel is a jumper's compatriot. On the other hand, Sandberg (2018), who used the term "indirect bias", and Zitzewitz (2006) found an opposite result for dressage and figure skating, respectively. We do not find evidence for compensating bias. Among other factors, this discrepancy might be explained by differences between our study and the previous studies in terms of dealing with home advantage in the analyses.

[4] For instance, Pope et al. (2018) performed a follow-up study to Price and Wolfers (2010) and showed that the racial bias among NBA referees disappeared after widespread media coverage.
[5] For example, the findings of Zitzewitz (2006) were summarized and discussed in the article "How ski jumping gets Olympic judging right (and figure skating gets it wrong)" by E. Zitzewitz, in the Washington Post on 12 February 2014 (https://www.washingtonpost.com/news/monkey-cage/wp/2014/02/12/how-ski-jumping-gets-olympic-judging-right-and-figure-skating-gets-it-wrong/).
[6] This result adds to the previous finding of Elaad et al. (2018), who showed that the more corrupt the country, the higher the probability that a team will achieve the desired result in order to avoid relegation in the last soccer game of a season. It also relates to Fisman and Miguel (2007), who found that United Nations diplomats living in New York who represent governments from very corrupt countries accumulated significantly more unpaid parking violations than their counterparts from less corrupt countries.

In fact, when controlling for the home variable, we find that the compensating bias loses most of its magnitude and becomes insignificant, both statistically and economically.[7]

The remainder of the paper is organized as follows. In Section 2, we describe the institutional settings of ski jumping competitions. In Sections 3 and 4, we present the data and descriptive statistics, and the empirical strategy, respectively. In Section 5, we present the baseline results, while we explore effect heterogeneity in Section 6. In Section 7, we compare our results with the results in other studies. We offer concluding remarks in Section 8.

Ski jumping rules

Ski jumping is a sport in which athletes ski down a track to generate speed and then jump from a ramp, with the aim of maximizing the length of the jump and the style points awarded by a judging panel. Three different hill sizes (HS) are used in professional ski jumping events: normal hills (HS 85–109 m), large hills (HS 110–184 m), and flying hills (HS 185 m and larger). Usually, 50 competitors jump in the first round. In flying hills, this number is reduced to 40. Based on the results of the first round, the top 30 jumpers advance to the second round. The winner of the competition is the jumper with the highest number of aggregate points achieved in both rounds.[8]

The aggregate point score is determined by the jumping distance and the style points. The jumping distance is an objective performance measure and quantified in intervals of 0.5 m. This distance is converted to a point value that contributes to the aggregate score. In addition, there is a subjective performance evaluation by a judging panel. The panel consists of five judges from five different countries, one of which is always the host nation. These judges award style points for the execution of the jump, landing, and outrun, based on predefined judging criteria for each part of the jump. Each judge awards a score of between 0 and 20 points, with intervals of 0.5. The lowest and highest scores are truncated to exclude extreme votes. The three remaining scores are summed up to the

[7] See Section 7 for a detailed discussion on differences between our findings and findings in ski jumping, figure skating (Zitzewitz, 2006), and dressage competitions (Sandberg, 2018).
[8] At World Cup competitions, the top 30 athletes receive World Cup points and prize money. For each World Cup point, the jumpers receive 100 Swiss francs (CHF), which amounts to 10,000 CHF for the winner of the competition. Extra prizes are awarded for special competitions such as the Four Hills Tournament (see FIS, 2017a, for additional information).

total style points. The athletes also receive compensation points for the starting gate and wind conditions to make the competition safer and fairer.

The judges of the panel are considered highly skilled and professional experts in this task. They are selected by the international governing body for winter sports, the Fédération Internationale de Ski (FIS). Judges must have a minimum of three years of experience at the national level, followed by a qualification period of at least two additional years. After the successful completion of the practical examination, the candidates receive their license to judge international ski jumping competitions. Moreover, ongoing training and an annual certification program is required to keep the status as an officially licensed judge (FIS, 2017b).

The judging process is designed to ensure the independent and discrete decision-making of the panel. According to the rules of the FIS (2017b), the athletes' performances must be evaluated objectively and without any prejudice. No communication with others is allowed and the decision must be entered into the scoring system without any assistance. Moreover, the judging tower where the judges are located is constructed in a way that provides optimal conditions for executing the judging task and ensuring compliance with the rules. More specifically, the tower is located at the side of the jumping hill such that each judge can clearly observe all parts of the jump. In addition, each judge has their own compartment in the judging tower so that they cannot view the scores of the other judges or be distracted by others.

Data and descriptive statistics

We collected data from the official website of the FIS on all men's World Cups, Nordic World Ski Championships, and Olympic Games (in Sochi 2014) for the seasons between 2010/11 and 2016/17. These are the most prestigious tournaments in professional ski jumping. This period was selected because of the introduction of wind and gate compensation points in the 2010/11 season.

For each jump, we have full information on athletes' names and nationality, competition date, and hill characteristics. We also have information on the judges' names and nationalities, as well as the individual judges' style point scores for each jump.

As summarized in Table 1, the data include performances of 268 jumpers from 24 countries, evaluated by 172 different judges from 19 countries, covering 203 competitions. Overall, we have information on 15,355 jumps, each of which was evaluated by five different judges, totaling up to 76,775 different jump evaluations. As described in Table A1 in the Appendix, the competitions took place in 14 countries. Events were most frequently held in Norway and Germany, with 38 competitions in each country. German judges were part of the panels in 71 percent of competitions, followed by Norwegian judges (59 percent) as the second most frequent country.

Table 1. Sample size

No. of ski jumpers	268
No. of ski jumper countries	24
No. of judges	172
No. of judge countries	19
No. of total competitions	203
No. of World Cups	165
No. of Four Hills	28
No. of Nordic World Championships	8
No. of Olympic Games	2
No. of jumps (performances)	15,355
Average no. of jumps per athlete	57.29 (81.70)
Average no. of jumps per athlete and season	17.73 (17.28)
No. of style point scores	76,775
Average no. of scores per judge	446.37 (291.87)
Average no. of scores per judge and season	143.24 (66.41)

Note: Standard deviations are presented in parentheses.

Table 2 provides the summary statistics for subsamples considering whether a judge and jumper are from the same country or not. On average, judges assign a higher score to their compatriots. In 9 percent of cases (6,941 evaluations overall), a judge was a compatriot of the evaluated jumper. This means that, in 36 percent of cases, four judges of the panel evaluated a compatriot of the remaining fifth judge of this panel. In addition, compatriot jumpers compete more frequently in their home countries and also perform better jumps in terms of jumping distance.

Table 2. Descriptive statistics

	Ski jumpers	
	Compatriot jumpers	Non-compatriot jumpers
Style points		
Mean	17.61	17.48
(overall SD)	1.05	1.07
(within-jump SD)		0.31
Min-max	5.0–20.0	4.0–20.0
Compatriot on panel		
Mean	0	0.40
Home event		
Mean	0.30	0.12
Jumping distance		
Mean	133.59	131.42
(overall SD)	31.43	30.24
Min-max	55.0–251.5	51.0–251.5
Wind points		
Mean	-0.92	-0.87
(overall SD)	8.30	8.44
Min-max	-34.9–43.4	-34.9–45.7
Gate points		
Mean	0.18	0.18
(overall SD)	4.56	4.53
Min-max	-29.4–45.2	-29.4–52.7
Country CPI score (2012–2017)		
Mean	72.90	70.82
(overall SD)	13.20	16.83
Min-max	28.33–88.67	28.33–88.67
No. of observations	6,941	69,834

Note: Standard deviations are presented only for metrical variables. CPI denotes the Corruption Perceptions Index published by Transparency International. Starting in 2012, the CPI uses a standardized scale from zero (very corrupt) to 100 (very uncorrupt) and includes information from several sources of the respective and previous years. For additional details on the CPI, see https://www.transparency.org/en/cpi (last accessed on October 16, 2020). Given a very small within-country CPI variation, we use the average CPI score for each country between the years 2012 and 2017.

Empirical strategy

In order to explore a possible nationalistic bias in professional ski jumping, we use *style points* awarded by each judge for a given jump as the unit of observation. In general, it is quite challenging to study the effect of a nationalistic bias on performance evaluation. Obviously, a naïve approach of correlating a dummy variable evaluating a compatriot jumper with the style points would yield biased and inconsistent estimates because a jumper's unobserved ability is likely to affect their performance, and therefore the decision-making of the judges. For example, it is possible that jumpers whose compatriot is on the panel have, on average, a higher quality, given that both the jumper and the judge come from nations where ski jumping is more popular. However, our data allow us to compare the style points of a compatriot judge with those of non-compatriot judges within

the same jump. In other words, we compare the scores from different judges who observed the same performance, estimating the following model,

$$style\ points_{jip} = \alpha_1\ compatriot\ jumper_{jip} + \theta_p + \lambda_{js} + \varepsilon_{jip} \quad (1)$$

where the dependent variable $style\ points_{jip}$ denotes the style points that judge j assigns to jumper i for jump p. The variable $compatriot\ jumper_{jip}$ is a dummy variable, equal to one if judge j and jumper i are from the same country; θ_p represents jump fixed effects. To control for idiosyncratic tendencies across judges (such as leniency or strictness), which may differ between judges, but also within a judge over the years, we use judge-per-season fixed effects, which is denoted by λ_{js}. A positive sign of α_1 implies a bias in favor of a compatriot jumper (in-group bias), whereas a negative sign of α_1 implies a bias against a compatriot jumper (out-group bias).

Beyond the issue of a nationalistic bias, another concern is that non-compatriot judges will assign lower (or higher) scores to jumpers if they have a compatriot judge on the panel (Zitzewitz, 2006; Sandberg, 2018). Obviously, any type of compensation (or reciprocation) is not allowed and might reinforce bias in evaluations by judging panels. To test the existence of a compensating bias, according to which judges consider whether one of the other judges is a compatriot of the evaluated jumper, we cannot use jump fixed effects because the composition of the judges is fixed within each jump. As noted earlier, a naïve approach of correlating a dummy variable of having a compatriot judge on the panel with the style points would yield biased and inconsistent estimates, because jumpers' unobserved ability is likely to affect their performance. However, ability can vary over time, differing over the years due to different preparations between seasons, injuries, or a natural decrease in physical strength that can appear at some point in a career. Hence, we need to take the different sources of unobserved heterogeneity into account. For example, Harb-Wu and Krumer (2019) investigated shooting accuracy in professional biathlon by using biathlete-per-season fixed effects.[9] As our panel data follow the same jumpers over many years, we follow the same approach as in Harb-Wu and Krumer

[9] In addition, see Genakos and Pagliero (2012) and Genakos et al. (2015) for a discussion about fixed effects estimations in multi-stage sports competitions.

(2019) and use jumper-per-season fixed effects as well as competition fixed effects, along with other observed characteristics of the jump, estimating the following model:

$$style\ points_{jipr} = \alpha_1\ compatriot\ jumper_{jipr} + \alpha_2\ compatriot\ on\ panel_{jipr} + \lambda_{js} + \delta_{is} + \mu_r + X_{ipr} + \varepsilon_{jipr} \quad (2)$$

Here, $compatriot\ on\ panel_{jipr}$ is a dummy variable that receives the value of one if judge j has a colleague on the judging panel of competition round r who is a compatriot of jumper i. This specification includes fixed effects for judges-per-season (λ_{js}), jumpers-per-season (δ_{is}), and each competition round (μ_r). X_{ipr} is our set of controls that includes a dummy variable for whether jumpers compete in their home country. It also includes an objective performance measure (i.e., the length of the jump), which is fully observed, and its squared term, as well as the wind and gate compensation points to observe the different conditions between jumps. These wind and gate compensation points, which were absent before 2010, enable us to better control for the objective quality of the jump. For this identification approach, we need to assume that there is no correlation between the composition of nationalities on the judging panel and the quality of jumps beyond what is already captured by the observables. A positive sign of α_2 implies bias in favor of jumpers who have a compatriot judge on the panel (positive reciprocation bias), while a negative sign of α_2 implies bias against such jumpers (negative compensating bias).

Baseline results

In Column 1 of Table 3, we present the results from model (1), controlling for jump fixed effects. Standard errors, which are three-way clustered at the judge, jumper, and jump level, appear in parentheses. We find that judges assign 0.09 style points more to their compatriot jumpers, corresponding to 29 percent of the within-jump standard deviation (as reported in Table 2). This result is statistically significant at the 1 percent level.[10]

To test the existence of a compensating bias, according to which judges take into account whether a certain jumper has a compatriot judge on the panel, we estimate model (2) because we cannot use jump fixed effects. First, we follow the approach of Zitzewitz

[10] A concern might be the possible risk of bias from censoring as there are observations with the maximal possible score of 20. However, we only observe 104 such observations (0.14 percent). Therefore, there is no serious risk of bias from censoring.

(2006) and Sandberg (2018), who did not use the dummy variable for whether jumpers compete in their home country (Column 2).[11] We find that the *compatriot on panel* variable is positive, but not statistically significant at conventional levels (p = 0.16), whereas the *compatriot jumper* coefficient increases slightly. However, because 30 percent of all jumps from a compatriot jumper in our sample were performed at a home event, we consider a potential home effect as highly relevant when analyzing performance evaluations. When additionally controlling for the home event (Column 3), the *compatriot on panel* variable loses most of its magnitude and becomes almost zero and highly insignificant (p = 0.86). In other words, having a counterpart on the judging panel who is from the same country as the jumper has no statistically significant effect on judges' evaluation.

Table 3. FE estimates for the judges' style point scores and the length of jump

Dependent variable	Style points			Length of jump
	(1)	(2)	(3)	(4)
Compatriot jumper	0.091***	0.109***	0.094***	
	(0.008)	(0.013)	(0.014)	
Compatriot on panel		0.018 (0.013)	0.002 (0.014)	0.072 (0.231)
Home event			0.056** (0.022)	1.860***
				(0.487)
Jump FE	Yes	No	No	No
Judge-per-season FE	Yes	Yes	Yes	No
Jumper-per-season FE	No	Yes	Yes	Yes
Competition-round FE	No	Yes	Yes	Yes
No. of observations	76,775	76,775	76,775	15,355

Note: In Columns 1–3, the dependent variable is the style points of each individual judge for a given jump. If no jump fixed effects are used, we control for performance indicators, which include the jumping distance and its squared term, as well as the wind and gate points. Standard errors are three-way clustered at the judge, jumper, and jump level and presented in parentheses. In Column 4, the dependent variable is the length of a jump in meters. Note that, in this case, the *compatriot on panel* variable defines whether one of the five judges is a compatriot. Here, we control for performance indicators, which include the wind and gate points. Standard errors are two-way clustered at the jumper and competition-round level and presented in parentheses. ***, **, * denote significance at the 1, 5, and 10 percent levels, respectively.

Theoretically, at least two explanations for such a home effect seem plausible. First, judges might be affected by the home crowd and thus bias their decision in favor of local jumpers (e.g., Garicano et al., 2005; Page and Page, 2010; Price et al., 2012; Waguespack and Salomon, 2015). Second, jumpers might simply perform better when competing in

[11] Although neither of the studies controlled for the home variable in that specification, they report in footnotes 11 (Zitzewitz, 2006) and 24 (Sandberg, 2018) that their findings on the existence of compensating bias are robust to exclusion of home participants.

their home country, resulting in higher style points. In order to test the latter, we explore whether jumpers make longer jumps when competing in their home country, estimating the following model:

$$length\ of\ jump_{ipr} = b_1\ home\ event_{ipr} + \alpha_2\ compatriot\ on\ panel_{ipr} + \delta_{is} + \mu_r + X_{ipr} + \varepsilon_{iPr} \quad (3)$$

Here, the dependent variable is the length of jump p in meters of jumper i in competition round r, $home\ event_{ipr}$ is a dummy variable that receives the value of one if a jumper competes in his home country. This specification includes fixed effects for jumper-per-season (δ_{is}), and for each competition round (μ_r), as well as a dummy of whether a jumper has a compatriot judge on the panel. X_{ipr} is our set of controls that includes the wind and gate compensation points.

In Column 4 of Table 3, we demonstrate that jumpers who compete in their home country jump, on average, 1.86 m longer compared with their jumps in competitions abroad. Similar to the case of subjective evaluation, having a compatriot judge on the panel has no statistically significant relationship with the length of jump, which is an objective measure of performance. Thus, we conclude that jumpers perform better when competing in their home country, which might also explain their higher style point scores. A possible explanation for such a home advantage is familiarity with the facilities (e.g., Barnett and Hilditch, 1993; Koning, 2011), a crucial factor in this technical discipline, which involves complex aerodynamic elements.

To test whether our findings on nationalistic bias are driven by extreme judges (outliers), we follow the approach of Sandberg (2018) by replacing the $compatriot\ jumper_{jip}$ variable in model (1) with an interaction term between a dummy for a specific judge and $compatriot\ jumper_{jip}$, and including judge fixed effects instead of judge-per-season fixed effects. We run this estimation separately for each of the 172 judges to obtain coefficients indicating how much, on average, judge j deviates from the other judges on the panel when jumper i is a compatriot minus how much, on average, judge j deviates from the other judges on the panel when jumper i is of another nationality.

In Figure 1, we present the results of this judge-specific degree of nationalistic bias. The figure shows that 76.7 percent of judges show a positive nationalistic bias and 49.1

percent are positive and statistically significant ($p < 0.05$), while only 2.5 percent of judges show a negative and statistically significant nationalistic bias.[12] Therefore, we conclude that the finding on positive nationalistic bias is not driven by only a few extremely biased judges.

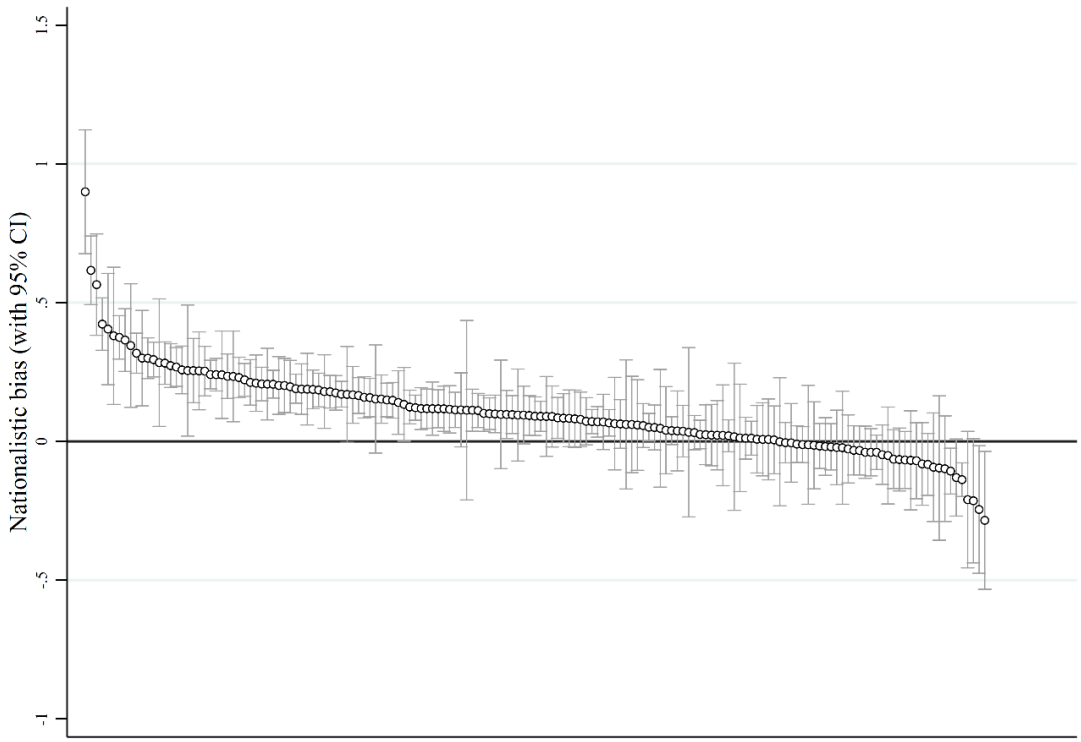

Figure 1. Judge-specific nationalistic bias

Notes: Each point represents a judge-specific estimate of the degree of nationalistic bias with 95 percent confidence intervals based on model (1) with judge fixed effects instead of judge-per-season fixed effects.

Likewise, to test whether our findings on the absence of compensating bias are driven by some abnormal patterns of individual judges, we present the judge-specific degree of compensating bias. For this analysis, we follow Sandberg (2018) to estimate a modified model (2), by replacing the $compatriot\ on\ panel_{jipr}$ variable with interaction terms between a dummy for each judge and $compatriot\ on\ panel_{jipr}$, and including judge fixed effects instead of judge-per-season fixed effects. In Figure 2, we present the results of this judge-specific degree of compensating bias. While 54.7 percent of judges show a positive compensating bias, only 8.1 percent are positive and statistically significant ($p <$

[12] We also estimated the judge-per-season specific degree of nationalistic bias. These results are available upon request and show a very similar pattern to that in Figure 1.

0.05). The remainder (45.3 percent) show a negative compensating bias,[13] with only 7.0 percent of judges showing a negative and statistically significant compensating bias. Taken together, the results suggest that compensating bias is not likely to play a significant role in performance evaluations by ski jumping judges.

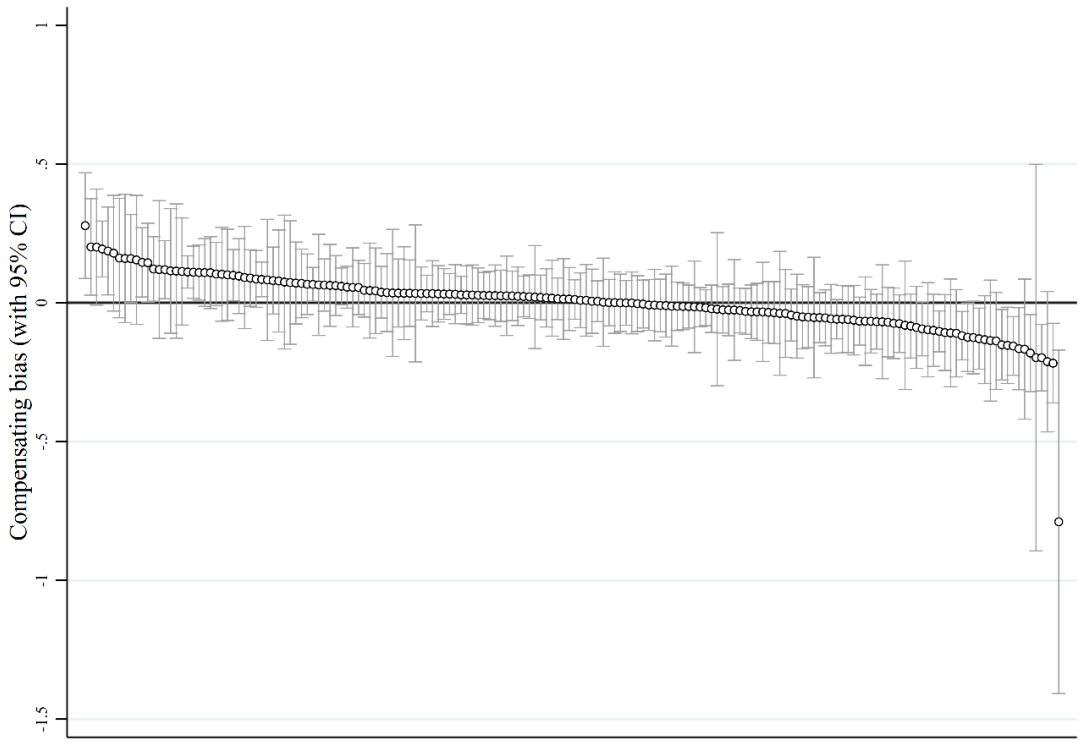

Figure 2. Judge-specific compensating bias

Notes: Each point represents a judge-specific estimate of the degree of compensating bias with 95 percent confidence intervals based on model (2) with judge fixed effects instead of judge-per-season fixed effects.

Effect heterogeneity

We further explore potential sources and variation of the nationalistic voting behavior of judges. First, we analyze event-specific variation of nationalistic bias. As ski jumping competitions consist of two rounds and only the top 30 jumpers qualify for the second round, their performances are decisive for determining the final ranking, including the winner and the distribution of prize money. Thus, stakes are higher in the second round and judges might have incentives to increase their nationalistic bias. Competitions also

[13] We also estimated the judge-per-season specific degree of compensating bias. These results are available upon request and show a very similar pattern to that in Figure 2.

have different hill size categories (normal, large, flying) and the importance of the style point score varies across these categories because of a different calculation of the final score. For example, in our data, the shares of style points from the final score are 45 percent, 44 percent, and only 30 percent for normal, large, and flying hills, respectively. Thus, the judges' contributions to the final outcome are less important at flying hill competitions, which reduces incentives for biased behavior. The nationalistic bias might also be stronger for events with a national character, such as the Olympic Games or World Championships, because national identity becomes more salient (e.g., Sandberg, 2018).[14]

In Table 4, we present the results of model (1) for the different subsamples of the data. Overall, the degree of nationalistic bias is similar for event rounds, hill sizes, and event types.[15] An exception is the Olympic Games, where the nationalistic bias is more than twice as large. However, when estimating model (1), including an interaction between $compatriot\ jumper$ and Olympic Games, we find no statistically significant difference (coefficient = 0.13, $p = 0.17$). Still, the nationalistic bias in the Olympic Games does not seem to be economically negligible, even if it is not statistically significant at conventional levels.

Table 4. Event-specific variation of nationalistic bias

Subsample estimations	No. of obs.	Coefficient	Standard error	*p*-Value
Round 1	49,020	0.095***	0.009	0.000
Round 1 Top 30	27,755	0.087***	0.009	0.000
Round 2	27,755	0.087***	0.010	0.000
Normal hills	6,215	0.089***	0.026	0.001
Large hills	58,435	0.093***	0.008	0.000
Flying hills	12,125	0.086***	0.016	0.000
World Cups	62,205	0.092***	0.008	0.000
Four Hills	10,610	0.088***	0.014	0.000
World Championships	3,185	0.082**	0.031	0.016
Olympic Games	775	0.217*	0.104	0.093

Note: The dependent variable is the style points of each individual judge for a given jump. All estimates are based on subsample estimations of model (1) with judge-per-season and jump fixed effects. Standard errors are three-way clustered at the judge, jumper, and jump level. ***, **, * denote significance at the 1, 5, and 10 percent levels, respectively.

[14] We also consider the Four Hills tournament as a separate event type because it includes the most prestigious World Cups in the calendar. The event has taken place in Germany and Austria each year since 1953. Winning all four events in one Four Hills Tournament edition is known as a grand slam. For additional information, see https://en.wikipedia.org/wiki/Four Hills Tournament.

[15] We also estimated model (1) with all data and interaction terms between $compatriot\ jumper$ and event round, hill size, and event type, finding no statistically significant differences, except for a slightly larger nationalistic bias in the first event round compared with the second round (coefficient = −0.02, p = 0.09).

We further test whether nationalistic bias might vary by country. In Figure 3, we present results for the nationalistic bias estimates of the 12 most observed countries in our dataset, based on subsample estimations of model (1) without judge-per-season fixed effects. We see that Russia has the highest nationalistic bias (0.22). Out of the 12 countries, Norway (0.00) and Finland (0.01) are the only two countries whose coefficients are negligibly small, both economically and statistically.[16] Such country-specific variation of nationalistic favoritism also seems a plausible explanation for the large but statistically insignificant effect for the Olympic Games. By looking at the composition of the judging panel at the 2014 Sochi Olympic Games, we find large differences in the judges' nationalistic bias. Again, Russia has the largest estimated bias (coefficient = 0.64, p = 0.01), which is 70 percent larger than the second highest estimated bias of Switzerland (coefficient = 0.38, p = 0.08).[17]

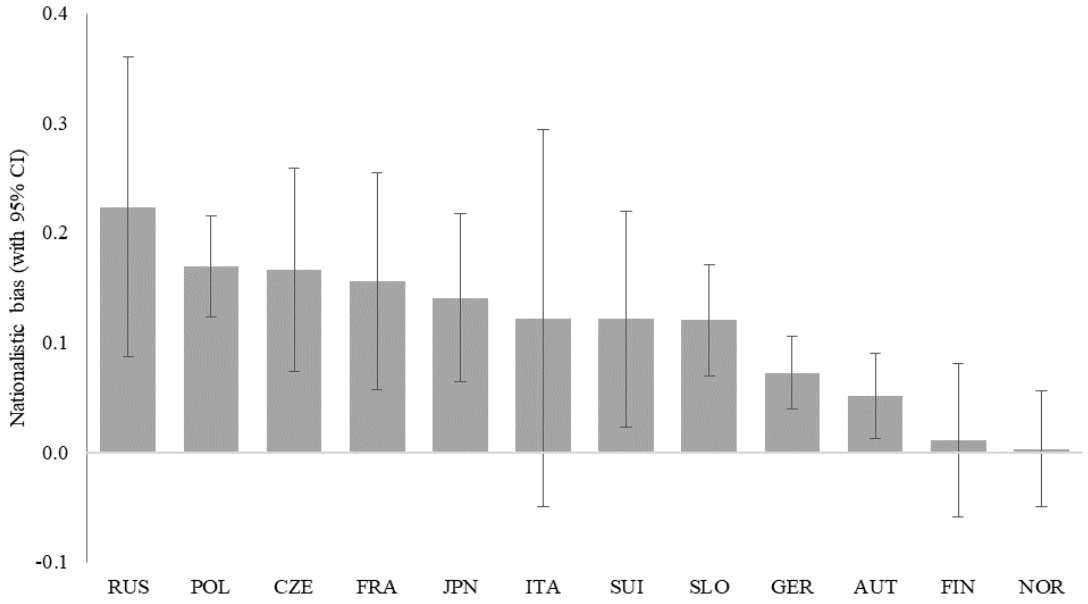

Figure 3. Country-specific variation of nationalistic bias

Notes: The figure shows the average nationalistic bias with 95 percent confidence intervals of judges when they evaluate performances of their compatriot jumpers. The estimates are based on subsample estimations of model (1) without judge-per-season fixed effects for the performances of all ski jumpers from the respective countries. The 12 countries are those with the most performance observations. The order of countries is based on the size of nationalistic bias. Please see Table A1 for the country abbreviations.

[16] Please note that Italian judges participated in only 21 percent of competitions compared to 59 percent and 53 percent of Norwegian and Finnish judges, respectively. For additional details, see Table A1 in the Appendix.

[17] The estimates are based on subsample estimations of model (1) for the 2014 Sochi Olympic Games and only for jumpers from countries whose judges were part of the panel. Because of data constraints, we could not use judge or judge-per-season fixed effects. The full set of results is available upon request.

In general, because the athletes' performances must be evaluated objectively and without any prejudice, such favoritism can be described as a corrupt type of behavior. Therefore, we explore whether nationalistic bias is related to the corruption perceptions index (CPI) of countries. In Figure 4, we demonstrate a negative relationship between the judge-specific nationalistic bias and the CPI score for a country.

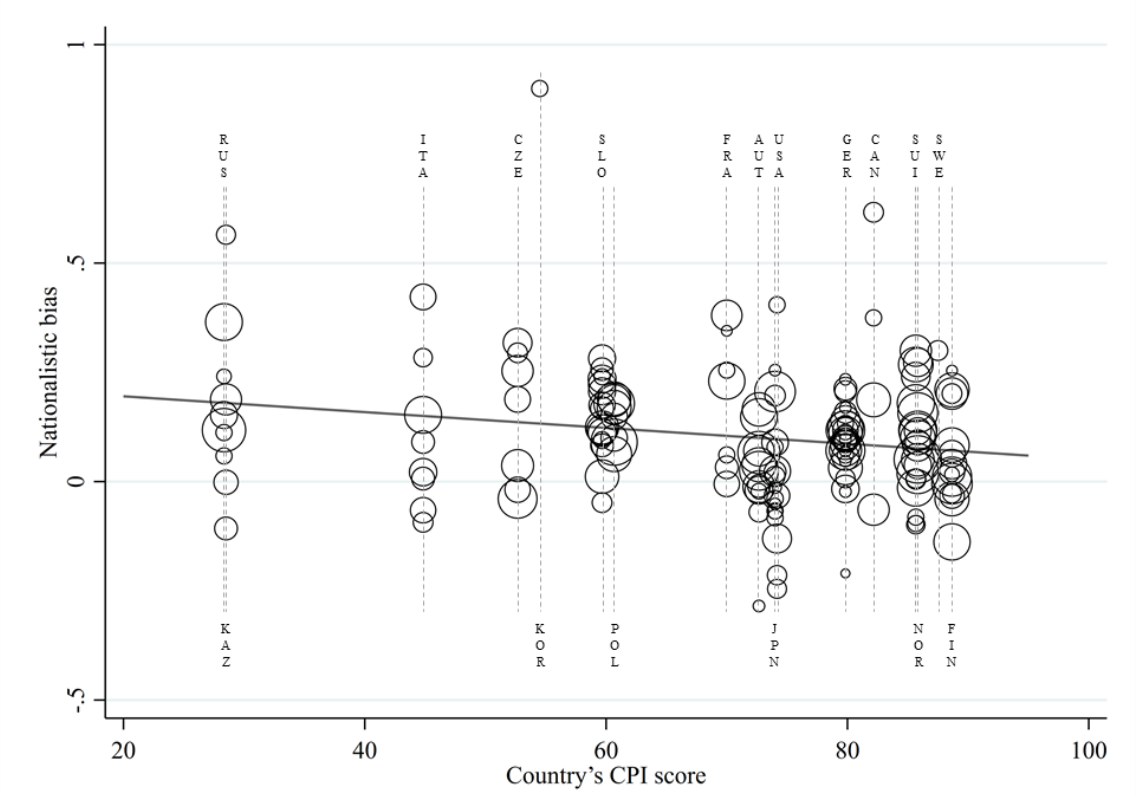

Figure 4. Nationalistic bias and the CPI

Notes: The circles show the judge-specific nationalistic bias. The size of the circles is relative to the number of observations for each judge in the data. The dashed vertical lines label the respective countries at the level of their CPI score. The regression line depicts the linear relationship between the judge-specific bias and the CPI score of the judges' countries.

The coefficient of the corresponding regression is −0.002 and it is statistically significant (p = 0.01).[18] In other words, the higher the CPI (less corrupt country), the lower the nationalistic bias. To put this result into perspective, an increase in one standard deviation

[18] The regression is based on model (1) and includes an interaction term between $compatriot\ jumper$ and the CPI score to estimate the relationship. We also run an alternative specification where we weigh by the number of observations per country. The results are very similar (coefficient = −0.002, p = 0.02).

in CPI reduces the nationalistic bias by 0.03 style points, which is 10 percent of the within-jump standard deviation of the evaluation of style points.

Finally, because Russia had the highest estimated nationalistic bias in the 2014 Olympic Games, but was also the only country that hosted Olympic Games in our data, it is possible that our findings on the relationship between the CPI and nationalistic bias are driven by hosting the Olympic Games and not by Russia per se. To obviate this concern, we remove the data of the Olympic Games and perform similar analyses to those in Figures 3 and 4. The results presented in Figures A1 and A2 in the Online Appendix show a very similar pattern. This finding is in line with previous cross-country evidence on positive relationships between unethical behavior and corruption levels in experimental settings (Barr and Serra, 2010; Gächter and Schulz, 2016) and non-experimental settings (Zitzewitz, 2006; Fisman and Miguel, 2007; Elaad et al., 2018).

A comparative view on nationalistic and compensating biases

Next, we compare the magnitude of nationalistic and compensating biases in our paper with the biases reported in Zitzewitz (2006) and Sandberg (2018).[19] Given the different scale of scores between the different sports, we present estimates standardized by the within-performance standard deviation. Figure 5 compares the standardized nationalistic bias estimates across studies. The nationalistic bias is the smallest in dressage (0.24), followed by our estimate for ski jumping (0.29). In comparison, the standardized coefficients for ski jumping (0.44) and figure skating (0.47), as reported in Zitzewitz (2006), are considerably larger.[20]

[19] Because Zitzewitz (2014), using a figure-skating setting, was unable to differentiate between nationalistic and compensating biases and Scholten et al. (2020), using a ski-jumping setting, employed a different estimation approach, neglecting some key issues, as mentioned in the Introduction, their results are hardly comparable with ours and, as such, are not considered here.

[20] We also find a similar pattern when we standardize the point estimates by the overall-performance standard deviation, which yields a standardized coefficient of 0.09 for our ski jumping estimations, 0.13 for both ski jumping and figure skating (Zitzewitz, 2006), and 0.07 for dressage (Sandberg, 2018).

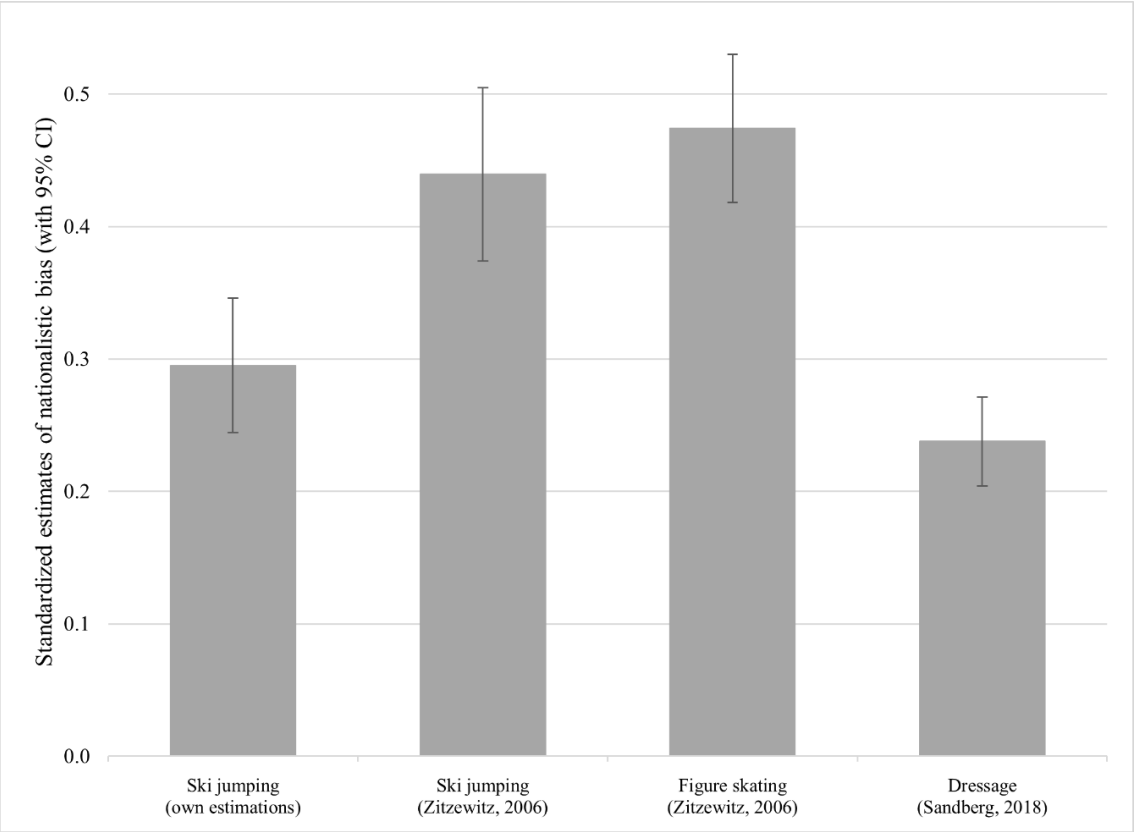

Figure 5. Standardized nationalistic bias estimates across studies

Notes: The figure shows the estimates of nationalistic bias with 95 percent confidence intervals, both standardized by the within-performance standard deviation. To calculate the values presented in this figure, we use the non-standardized point estimates from Column 1 in Table 3 for ski jumping (our own estimations); we also use values for ski jumping and figure skating from Zitzewitz (2006, Table 3, Lines 5 and 3, respectively), and values for dressage from Sandberg (2018, Table 3, Column 1).

When comparing both findings for ski jumping, it should be noted that Zitzewitz (2006) only used data from 2002, an Olympic year. In both studies related to ski jumping, the nationalistic bias in the Olympic Games is the highest: 0.26 in Zitzewitz (2006)[21] and 0.22 in our study. As such, the inclusion of the Olympic Games generally increases the average estimate of nationalistic bias in ski jumping. However, in our case, the share of the Olympic Games is only 1 percent of the overall number of observations, while it is 13 percent in Zitzewitz (2006). When comparing our finding for ski jumping with the finding for figure skating, it should be noted that ice dancing accounts for one-third of the data on figure skating in Zitzewitz (2006). As noted by the author, "biases are larger where

[21] See Panel A, Line 2 in Table 5 of Zitzewitz (2006).

scoring is more subjective, as it is for ice dancing, where skaters do not have as many mandatory deductions for falls, and for artistic impression as opposed to technical merit scores" (Zitzewitz, 2006, p. 79). This is in line with a recent paper by Joustra et al. (2021), who found a significant advantage for later performances in female gymnastics, which is likely to be driven by the existence of subjective evaluation only in female competitions considering artistry. In fact, the nationalistic bias found by Zitzewitz (2006) for ice dancing is 33 percent higher than that for non-ice dancing disciplines.[22]

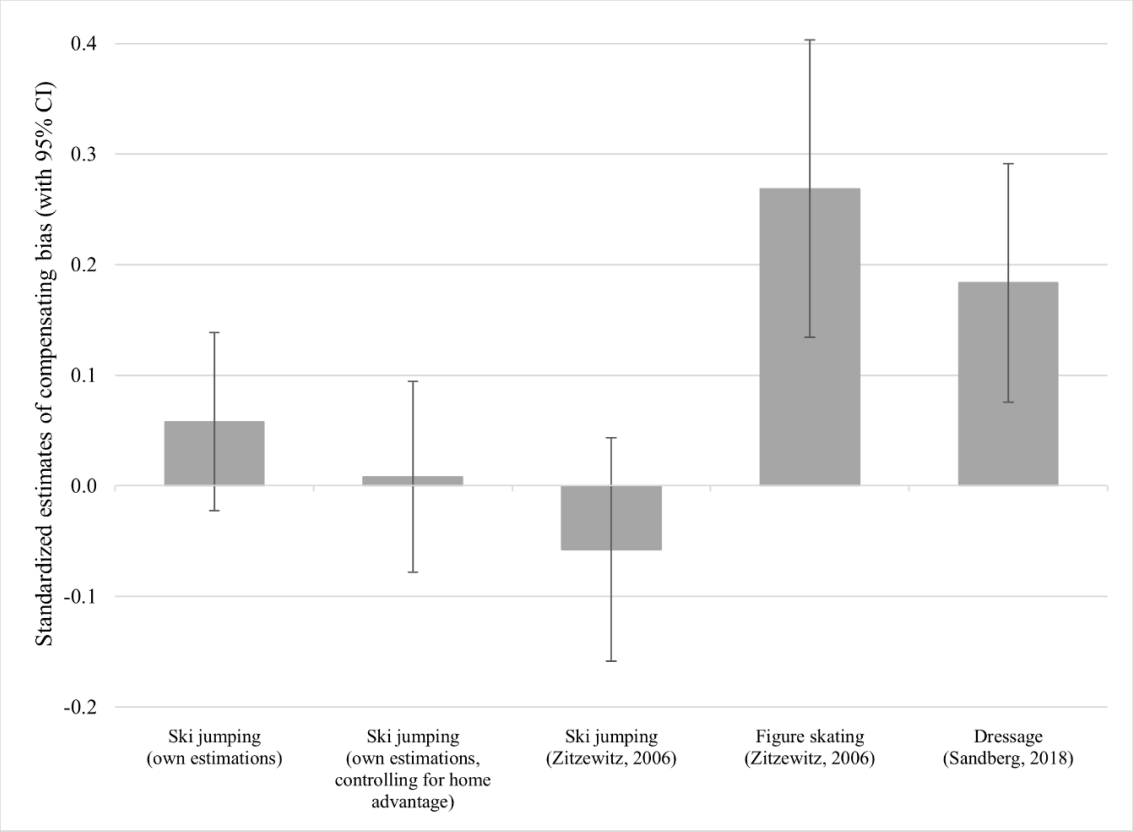

Figure 6. Standardized compensating bias estimates across studies

Notes: The figure shows the estimates of compensating bias with 95 percent confidence intervals, both standardized by the within-performance standard deviation. To calculate the values presented in this figure, we use the non-standardized point estimates from Column 2 in Table 3 for ski jumping (our own estimations) and from Column 3 in Table 3 for ski jumping (our own estimations), where we also control for home advantage; we also use values for ski jumping and figure skating from Zitzewitz (2006, Table 4, Panel A/Line 5b and Panel B/Line 5, respectively), and for dressage from Sandberg (2018, Table 5, Column 2).

[22] See Panel B in Table 5 of Zitzewitz (2006).

Figure 6 compares the standardized compensating bias estimates across studies. While both of our standardized coefficients (i.e., with and without controlling for home advantage) are positive and the standardized coefficient for ski jumping in Zitzewitz (2006) is negative, none of these differs significantly from zero. In contrast to these findings, the standardized coefficients for figure skating (0.27) and dressage (0.18) are comparably large and significant.[23] One possible reason for this might be that the estimating equations for figure skating and dressage do not include the home variable. However, the home variable is also excluded from the estimating equation for ski jumping in Zitzewitz (2006), and both Sandberg (2018) and Zitzewitz (2006) have reported that their results are robust to exclusion of home participants.

Another possible reason why we observe positive reciprocation biases in figure skating and dressage in contrast to ski jumping might be the differences in institutional settings that relate to truncation and exposure of scores in the sports. In the scoring system used in figure skating, judges' scores are transformed into votes about the skaters' relative performance. According to Zitzewitz (2006), such a system makes it easier to detect a defection from reciprocal arrangements than a system with continuous scores. Therefore, the transformation of scores into votes can make reciprocal arrangements easier to sustain. In dressage, a rule promotes consistency in scoring. According to this rule, the panel members must have an evaluation meeting after the competition if the scores for a performance differ by more than 5 percent among the judges. Thus, it seems possible that experienced dressage judges anticipate the nationalistic bias of their panel members and act accordingly – that is, they give better (and biased) scores to ensure consistency. This is different in ski jumping, where the truncation mechanism seems to lower incentives either for compensation (because extreme votes are excluded) or for consistency (because there is no such 5 percent rule).

[23] We also find a similar pattern when we standardize the point estimates of compensating bias by the overall-performance standard deviation, which yields a standardized coefficient of 0.02 for our ski jumping estimations and 0.00 when controlling for home advantage, −0.03 for ski jumping and 0.07 for figure skating in Zitzewitz (2006), and 0.06 for dressage in Sandberg (2018).

Conclusion

In this paper, we have examined nationalistic bias in subjective evaluations by international experts, which has been shown to be a significant factor in previous studies (Zitzewitz, 2006; Sandberg, 2018). Our efforts in this regard are in line with the increasing importance of replication studies (Ioannidis and Doucouliagos, 2013; Open Science Collaboration, 2015) and crowdsourced research (Silberzahn and Uhlmann, 2015; Silberzahn et al., 2018).

Our findings confirm the existence of nationalistic voting of judges in professional ski jumping competitions more than a decade after this bias was first illustrated in similar settings. This nationalistic bias is found for a large share of judges and is positively associated with the level of corruption according to the CPI. Our results suggest that in-group favoritism is a strong feature of human behavior, especially in countries with a high prevalence of corruption in their institutional environment. In addition, unlike previous findings, our results show no evidence of strategic voting, according to which judges assign different scores to jumpers whose compatriots are present on the judging panel. This discrepancy can be partly explained by different approaches in dealing with home advantage and different institutional settings. Unlike figure skating and dressage, ski jumping uses a truncation mechanism, according to which the highest and lowest scores are excluded, which seems to lower incentives for strategic voting.

It is important to note that our results were obtained from fully observable sports competitions. Such in-group favoritism might even be stronger in less transparent settings that involve subjective decision-making, such as policymaking processes, judging in legal proceedings, human resource management, etc.

Finally, we call for future research to investigate nationalistic favoritism in other settings to create higher awareness of this primitive human instinct that has not yet disappeared. This call is particularly important during times when the entire humanity faces difficulties, such as COVID-19, where the immediate and natural desire is to protect in-group members, which could lead to an increased nationalistic favoritism.

References

Andrews, T. J., Smith, R. K., Hoggart, R. L., Ulrich, P. I., and Gouws, A. D. (2019), Neural correlates of group bias during natural viewing, *Cerebral Cortex 29*, 3380–3389.

Barnett, V. and Hilditch, S. (1993), The effect of an artificial pitch surface on home team performance in football (soccer), *Journal of the Royal Statistical Society: Series A (Statistics in Society) 156*, 39–50.

Barr, A. and Serra, D. (2010), Corruption and culture: an experimental analysis, *Journal of Public Economics 94*, 862–869.

Efferson, C., Lalive, R., and Fehr, E. (2008), The coevolution of cultural groups and ingroup favoritism, *Science 321*, 1844–1849.

Elaad, G., Krumer, A., and Kantor, J. (2018), Corruption and sensitive soccer games: cross- country evidence, *Journal of Law, Economics, and Organization 34*, 364–394.

Faltings, R., Krumer, A., and Lechner, M. (2021), Rot-jaune-verde. language and favoritism: evidence from swiss soccer, School of Economics and Political Science, Department of Economics, University of St Gallen, Working Paper.

Fédération Internationale de Ski (FIS) (2017a), *Rules for the FIS Ski Jumping World Cup (men)*, Edition 2017/2018, Oberhofen, Switzerland.

Fédération Internationale de Ski (FIS) (2017b), *The International Ski Competition Rules (ICR), Book III, Ski Jumping*, Oberhofen, Switzerland.

Fisman, R. and Miguel, E. (2007), Corruption, norms, and legal enforcement: evidence from diplomatic parking tickets, *Journal of Political Economy 115*, 1020–1048.

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

Gächter, S. and Schulz, J. F. (2016), Intrinsic honesty and the prevalence of rule violations across societies, *Nature 531*, 496–499.

Genakos, C. and Pagliero, M. (2012), Interim rank, risk taking, and performance in dynamic tournaments, *Journal of Political Economy 120*, 782–813.

Genakos, C., Pagliero, M., and Garbi, E. (2015), When pressure sinks performance: evidence from diving competitions, *Economics Letters 132*, 5–8.

Harb-Wu, K. and Krumer, A. (2019), Choking under pressure in front of a supportive audience: evidence from professional biathlon, *Journal of Economic Behavior & Organization 166*, 246–262.

Ioannidis, J. and Doucouliagos, C. (2013), What's to know about the credibility of empirical economics?, *Journal of Economic Surveys 27*, 997–1004.

Joustra, S. J., Koning, R. H., and Krumer, A. (2021), Order effects in elite gymnastics, *De Economist 169*, 21–35.

Koning, R. H. (2011), Home advantage in professional tennis, *Journal of Sports Sciences 29*, 19–27.

Open Science Collaboration (2015), Estimating the reproducibility of psychological science, *Science 349*, aac4716.

Page, K. and Page, L. (2010), Alone against the crowd: individual differences in referees' ability to cope under pressure, *Journal of Economic Psychology 31*, 192–199.

Pope, B. R. and Pope, N. G. (2015), Own-nationality bias: evidence from UEFA Champions League football referees, *Economic Inquiry 53*, 1292–1304.

Pope, D. G., Price, J., and Wolfers, J. (2018), Awareness reduces racial bias, *Management Science 64*, 4988–4995.

Price, J., Remer, M., and Stone, D. F. (2012), Subperfect game: profitable biases of NBA referees, *Journal of Economics & Management Strategy 21*, 271–300.

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

Sandberg, A. (2018), Competing identities: a field study of in-group bias among professional evaluators, *Economic Journal 128*, 2131–2159.

Scholten, H., Schneemann, S., and 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*, 708–734.

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

Silberzahn, R. and Uhlmann, E. L. (2015), Crowdsourced research: many hands make tight work, *Nature News 526*, 189–191.

Silberzahn, R. et al. (2018), Many analysts, one data set: making transparent how variations in analytic choices affect results, *Advances in Methods and Practices in Psychological Science 1*, 337–356.

Spierdijk, L. and Vellekoop, M. (2009), The structure of bias in peer voting systems: lessons from the Eurovision Song Contest, *Empirical Economics 36*, 403–425.

Sumner, W. G. (1907), *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*, Ginn and Company, Boston, MA.

Waguespack, D. M. and Salomon, R. (2015), Quality, subjectivity, and sustained superior performance at the Olympic Games, *Management Science 62*, 286–300.

Yuki, M. (2003), Intergroup comparison versus intragroup relationships: a cross-cultural examination of social identity theory in North American and East Asian cultural contexts, *Social Psychology Quarterly 66*, 166–183.

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

Zitzewitz, E. (2014), Does transparency reduce favoritism and corruption? Evidence from the reform of figure skating judging, *Journal of Sports Economics 15*, 3–30.

Appendix

Table A1. Frequencies of countries by groups of jumpers, judges, and competitions

Country name	Country code	Jumpers	Jumps	Judges	Competitions	Judges in competitions
Austria	AUT	27	2077	12	19	91 (45%)
Bulgaria	BUL	1	110	0	0	0 (0%)
Canada	CAN	5	113	4	0	22 (11%)
Czech Republic	CZE	15	1171	7	6	50 (25%)
Estonia	EST	4	70	0	0	0 (0%)
Finland	FIN	19	717	16	23	107 (53%)
France	FRA	6	316	7	0	33 (16%)
Germany	GER	24	2098	30	38	145 (71%)
Greece	GRE	1	3	0	0	0 (0%)
Italy	ITA	8	299	8	4	43 (21%)
Japan	JPN	27	1408	15	14	54 (27%)
Kazakhstan	KAZ	8	53	4	2	20 (10%)
South Korea	KOR	4	39	1	2	2 (1%)
Netherlands	NED	1	1	0	0	0 (0%)
Norway	NOR	27	2037	12	38	120 (59%)
Poland	POL	20	1666	8	16	79 (39%)
Romania	ROU	2	3	4	0	8 (4%)
Russia	RUS	18	688	6	8	38 (19%)
Slovenia	SLO	27	1774	15	15	84 (41%)
Switzerland	SUI	10	564	10	15	60 (30%)
Slovakia	SVK	1	2	2	0	14 (7%)
Sweden	SWE	2	10	4	3	22 (11%)
Ukraine	UKR	2	2	0	0	0 (0%)
USA	USA	9	134	7	0	23 (11%)
Total	24	268	15,355	172	203 competitions	

Note: The last column states the number of competitions in which the respective country has a judge on the panel. There are five judges in each competition. This is also presented as percentage share based on the total number of competitions in parentheses.

Appendix (online appendix)

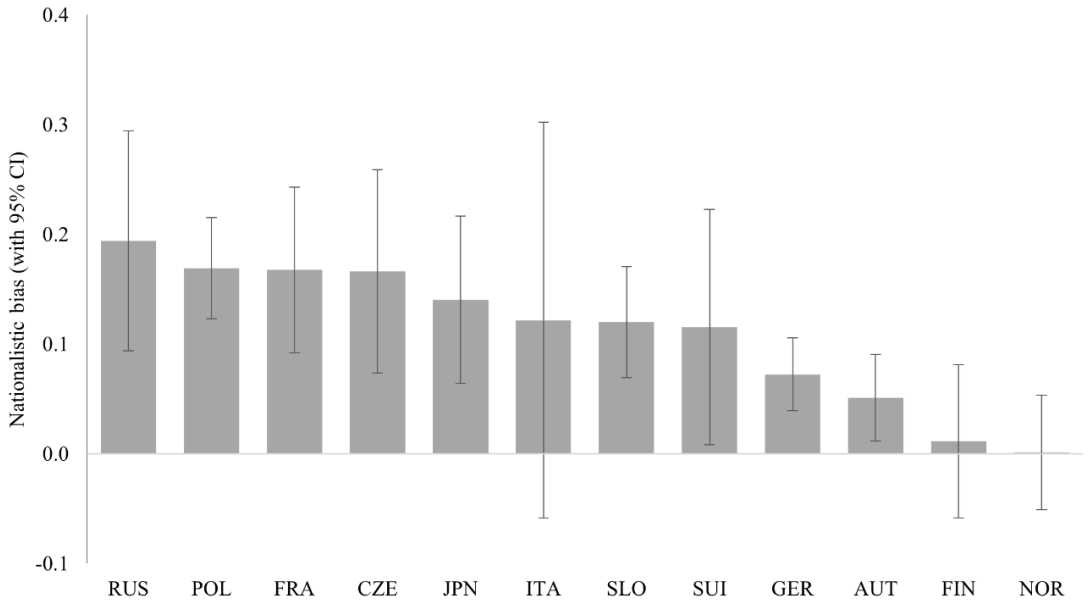

Figure A1. Country-specific variation of nationalistic bias (subsample excluding Olympic Games)

Notes: The figure shows the average nationalistic bias with 95 percent confidence intervals of judges when they evaluate performances of their compatriot jumpers. The estimates are based on subsample estimations of model (1) without judge-per-season fixed effects for the performances of all ski jumpers from the respective countries, excluding the Olympic Games. The 12 countries are those with the most performance observations. The order of countries is based on the size of nationalistic bias.

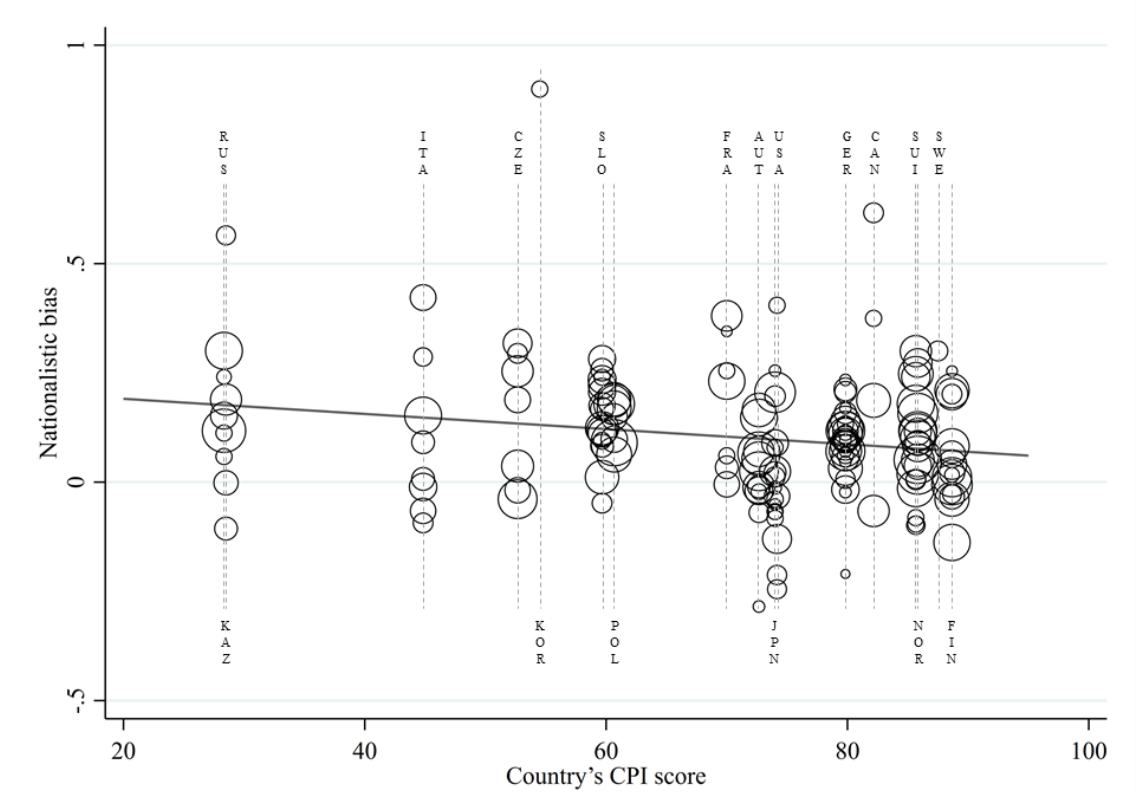

Figure A2. Nationalistic bias and the Corruption Perceptions Index (CPI) (subsample excluding Olympic Games)

Notes: The circles show the judge-specific nationalistic bias. The size of the circles is relative to the number of observations for each judge in the data. The dashed vertical lines label the respective countries at the level of their CPI score. The regression line depicts the linear relationship between the judge-specific bias and the CPI score of the judges' countries.

5 Discussion

The empirical studies presented in the previous Chapter 4 provide valuable insights into the role of fairness and identity concerns of spectators and judges in professional sports. In the following, the empirical findings and theoretical implications are summarized and discussed (Subchapter 5.1). This is followed by a discussion of the practical implications for sports policy and management (Subchapter 5.2).

5.1 Empirical findings and theoretical implications

Study 1 and Study 2 of this dissertation focus on fairness and national identity concerns of sport spectators. Both studies rely on individual-level survey data and meaningful sports demand settings to test the relations of interest. Study 3 focuses on national identity concerns of sport judges and uses data on performance scores awarded by ski jumping judges at the most important international competitions. Overall, the studies suggest that national identity concerns are important for spectator and judge behavior, while fairness concerns seem less important for spectator behavior, as discussed in more detail in the following sections (5.1.1–5.1.3).

5.1.1 Trust in fairness and spectator behavior

Study 1 focuses on fairness concerns of sport spectators in the context of a doping scandal and the TV demand of international track and field events. The first part tests the relation between trust in fairness and the revealed TV consumption of the EAC, finding that trust in fairness has no significant impact on TV consumption. This finding generally questions the relevance of trust in fairness for spectator demand. The second part focuses on a real doping case and a sports event scenario. The results show that awareness of past doping scandals of a popular athlete negatively affects the trust in athletes' fair conduct and integrity, but there is no impact on TV demand. Study 1 thus provides empirical evidence that a major doping case can actually lead to a long-term loss of spectators' trust in fairness, but it does not support the relevance of trust in fairness either for spectator demand in general or after a major doping scandal.

Accordingly, these findings generally question the role of fairness concerns for spectator demand and are thus in line with suggestions by Buraimo et al. (2016) and Manoli et al.

(2020). Even if spectators' trust in fairness is reduced after past doping scandals, it does not necessarily reduce spectator demand, suggesting that athletes' fair conduct is not, or at least less, relevant for following sports events. One possible explanation for this finding is that spectators derive utility from witnessing scandals and related media coverage, meaning that doping cases themselves provide some benefits, as argued by Van Reeth (2013). An alternative explanation is that the (long-term) costs of watching a possibly unfair competition do not outweigh the overall benefits of sports entertainment. This interpretation is in line with previous findings in baseball, where doping news only has a short-term effect on attendance and TV demand and where doping news from opposing teams has no effect on local TV demand (Cisyk, 2020; Cisyk & Courty, 2017). Here, the spectators' response to doping cases only seems to be a temporary expression of discontent. Study 1 thus emphasizes that the doping–demand relation and its underlying mechanisms are complex and may depend on several other aspects like, for instance, type of scandal and sport, or athlete characteristics (e.g., superstar, compatriot, or athlete from favorite team or not). Nonetheless, popular athletes convicted of doping seem to act as distrust ambassadors, promoting the unfairness of the competition to the spectators.

Furthermore, previous research emphasized the relevance of trust in fairness of the financial system for investor behavior (Guiso et al., 2008). This may suggest that fairness concerns are more relevant if possible unfairness entails higher (personal) costs (e.g., lower financial wealth versus less enjoyment in watching sports). Moreover, similar to previous studies, the findings of Study 1 show that scandals reduce trust in the involved players, but in contrast to these studies, trust does not affect demand (Ding et al., 2013; Giannetti & Wang, 2016). This again suggests that the importance of trust in fairness for behavior on the demand side may depend on setting-specific aspects (e.g., market characteristics or product type).

5.1.2 National identity and spectator behavior

Furthermore, Study 1 and Study 2 provide insights on the role of national identity concerns for spectator demand. In Study 1, the results of the zero-inflated Poisson regressions indicate that national identity concerns, measured by the level of patriotism, are positively related to the frequency of the TV consumption of the EAC. This finding indicates that spectators' national identity affects their demand for *international* sports. Study 2 focuses on national identity concerns in the context of countries' political

relations and on-site spectator demand. The study tests the relation between the perceived friendliness toward the COO and the WTP for admission to relocated European cup finals and league games in a hypothetical scenario. The results show a higher willingness-to-pay for admission to the relocated games if the COO of the competing teams is perceived as friendly. This relation is also confirmed by an IV approach, in which the instrument proxies the attitude toward foreign countries by the level of national identity in the residential area.

Study 2 thus provides empirical evidence that the political relations between the importing country and the COO impact spectator demand of popular *domestic* sports. From a theoretical perspective, the findings suggest that the perceived friendliness toward the COO seem to make similarities or differences between the home country and COO salient, which is used by spectators for national identity formation, affecting their demand for sports. The findings of Study 1 and Study 2 are both in line with social identity theory (Tajfel & Turner, 1986) and identity economics (Akerlof & Kranton, 2000), and support the relevance of national identity concerns for the demand of imported consumer goods.

Both studies also contribute to the limited literature on spectators' national identity concerns and sports demand. Study 1 confirms previous findings, confirming that national identity concerns drive TV demand for international sports competitions (e.g., Konjer et al., 2017; Nüesch & Franck, 2009). The findings of Study 2 are in line with the findings of the general COO and consumer boycott literature (e.g., Pandya & Venkatesan 2016; Verlegh & Steenkamp, 1999), and imply that the perceived status of the political relations between importing country and COO impact consumer behavior and demand even in the absence of severe political conflicts. This supports the reasoning of Riefler and Diamantopoulos (2007) that this issue seems particularly relevant for products that are non-durable and easy to substitute.

5.1.3 National identity and judge behavior

Study 3 also focuses on national identity concerns but in the context of the judging behavior of professional experts and sports performance evaluations. The study first tests the presence of nationalistic bias among judges in international ski jumping competitions over several seasons. The results show that a large share of judges favors their compatriot athletes by assigning them higher style point scores. Therefore, nationalistic bias still

exists in ski jumping a decade after this issue was first identified and raised by Zitzewitz (2006). It should therefore be regarded as a prevalent issue in subjective performance evaluations. The results further show that nationalistic bias remains consistently strong across rounds and events, suggesting that neither variation in competition stakes nor in the salience of national identity alters in-group favoritism. The results also show that judges do not give, on average, different scores if another panel member and the ski jumper share the same nationality. Comparisons across sports further indicate that this seems to partly relate to the judging systems employed by the different sports. Besides this, nationalistic bias varies strongly by country and is higher in more corrupt countries, according to the CPI.

Overall, the findings suggest that a large share of ski jumping judges cannot resist their inherent preferences toward compatriot athletes in international competitions, especially if they come from countries with more corruption in their institutional environment. This suggests that nationalistic in-group favoritism is a strong and prevalent feature in behavior of professional experts and again supports the relevance of identity-based preferences in economic behavior (Akerlof & Kranton, 2000).

Based on these findings, Study 3 contributes to the literature on in-group favoritism in international sports, confirming the prevalence of nationalistic bias in performance evaluations (e.g., Emerson et al., 2009; Zitzewitz, 2006, 2014). However, it provides no evidence for strategic voting or that national salience creates temporary group identity for non-compatriot judges, as argued by Sandberg (2018). The differences in findings across sports seem to partly relate to the different judging systems, as already suggested by Zitzewitz (2006). Moreover, Study 3 contributes to the general literature on identity-based judging bias in subjective evaluations, confirming previous findings of in-group favoritism based on origin, i.e., ethnicity and race in national (e.g., Dee, 2005; Shayo & Zussman, 2011) and nationality in international settings (Feld et al., 2016).

All in all, the findings of the three studies of this dissertation provide empirical evidence that national identity concerns affect spectator and judge behavior in professional sports. The findings are consistent with previous research that focused on identity concerns in settings other than sports (i.e., studies on consumer animosity and boycott calls as well as studies on judging bias of professional experts) and provide further insights in this regard.

In particular, they shed further light on the question of which social group people identify themselves with when diverse social and cultural backgrounds are salient. In international contexts, economic agents seem to choose their social identity based on their national origin, likely due to a relatively high group status and / or the salience of national attributes.

5.2 Practical implications

Understanding the role of trust in fairness and national identity for spectator and judge behavior is of considerable interest for sports policy makers and managers because changes and drifts in spectator demand or performance evaluations can have serious financial consequences for sports organizers and competitors.

The findings of Study 1 on spectators' trust in fairness suggests that even a loss of trust driven by major doping scandals does not need to be of concern for sports organizers because it may not have negative long-term effects on spectator demand. Fans do not seem to completely refrain from following their favorite sports. However, since a major doping case of a popular athlete can reduce spectators' trust in fairness, it should be considered a serious issue from a sports policy perspective if professional sports is used to promote social values of fairness in order to justify public funding. It may also prevent sponsorship engagements if sponsors aim to benefit from an image transfer of positive fairness values.

Besides this, Study 1 and Study 2 emphasize the importance of national identity concerns for spectator demand. Findings of Study 1 suggest that sports managers and marketers may focus on spectators' national identify formation to promote TV broadcasts of international competitions. Since spectators with a higher level of patriotism also show a higher frequency of TV consumption, emphasizing national features of the sports competition in advertising and marketing may promote national identity formation, increasing patriotic sentiment and consequently demand. Furthermore, findings of Study 2 suggest that the political relations between the importing and the exporting countries are important for spectator demand when domestic sports games are relocated to foreign markets. As such, the spectators' political opinions about the COO may constitute a crucial component for sports leagues' expansion strategies and should thus be considered

by sports managers. For instance, if the COO is perceived as rather unfriendly in a target market (e.g., at the regional or country level), disregarding national attributes of the sports in advertising and marketing may reduce the impact of national identity concerns on spectator demand. Moreover, considering differences in the political opinions of target groups and the level of national identity across geographic markets seems advisable for developing internationalization strategies and target market selection in the first place.

In addition, Study 3 further emphasizes the importance of national identity concerns for sports judge behavior. The findings indicate that judges favor their compatriot athletes in their performance evaluations. This nationalistic bias was found to be prevalent for a large share of judges and across several seasons, events, and competitions rounds. This is a serious issue from a sports policy and management perspective: First, sports organizers are in charge to set and enforce the rules of sports competitions and, therefore, appoint and pay judges to act in the organizers' interest, i.e., to be impartial in their evaluations. Second, systematic judging bias may have serious consequences for sports competition outcomes, affecting contest results and prize money distribution. Therefore, sports organizers should take the possibility of nationalistic bias of judges in international sports competitions seriously and should develop and implement judging systems and measures in order to remove in-group bias of individual judges in judging panel evaluations.

6 Conclusion

This dissertation focuses on the role of social concerns for economic behavior in the context of professional sports. In this regard, the three empirical studies conducted within the scope of this dissertation address five previously identified research gaps and desiderata, which relate to fairness and national identity concerns and sports spectator and judge behavior. The first study explores the impact of trust in fairness on spectators' TV demand and its role in the doping–demand relation; the second study explores the impact of countries' political relations on on-site demand; and the third study explores judges' nationalistic bias in sports performance evaluations.

Summarizing the findings of the three studies, there is no support for the importance of fairness concerns for spectator demand but consistent support for the importance of national identity concerns for spectator and judge behavior in professional sports. More precisely, while the first study does not find that trust in fairness matters for spectator behavior, all three studies find evidence that national identity concerns affect spectator and judge behavior. In this way, the dissertation attempts to contribute to the sports economics literature by emphasizing the need to consider social (identity) concerns in professional sports settings. This may become even more relevant in the future because of the increasingly international scope of top-level sports competitions and the accompanying sports industry.

Although the studies of this dissertation provide valuable insights into the role of social concerns in professional sports, they also have some limitations. Some major shortcomings of the empirical research are discussed in the following, including opportunities for future research.

The first and second study rely on carefully designed hypothetical scenarios and a stated preference approach to proxy spectator demand. Although this approach allows the use of individual-level data to directly test the relations of interest, the hypothetical context also entails some shortcomings, i.e., the selection of the sports events and the elicitation designs. Another limitation refers to the impact of trust in fairness on spectator demand. Fairness concerns may be particularly important for spectator behavior in the immediate period after a doping scandal. We actually intended to exploit a real doping case as an

exogenous shock in our panel survey; however, we could not identify any serious anti-doping rule violations or similar illegal conduct between the two survey waves of the first study. With regard to both limitations, it would be advisable for future research to test the impact of spectators' trust in fairness after a real doping shock as well as the status of political relations on spectator demand with *actual* consumer choice data.

Furthermore, the first and second studies are limited with regard to exploring the scope of national identity concerns for spectator demand. Each study focused on one particular country (Germany and the US) as the setting in their analyses, but national identity may shape individual preferences differently across countries.[19] Therefore, future research may also consider potential cross-country differences in national identity concerns and spectator preferences.

Regarding the role of national identity for judge behavior, the third study explores and discusses various potential sources and variation of nationalistic bias in judges' performance evaluations in ski jumping but also in comparison to figure skating and dressage. Although some of the reported differences in the findings seem to relate to the judging systems employed by the different sports and the corresponding incentive structure, the empirical analysis does not allow a direct test on how measures, such as the truncation of extreme scores, rules for consistency in scoring, or transparency rules, mitigate nationalistic bias of sports judges. Future research should thus focus on the judging systems' measures that are used in sports or other organizations to mitigate judging biases.

[19] For instance, see Mayda and Rodrik (2005) for cross-country differences in national identity and how it correlates to individual preferences for international trade.

References

Abeza, G., O'Reilly, N., Prior, D., Huybers, T., & Mazanov, J. (2020). The impact of scandal on sport consumption: Do different scandal types have different levels of influence on different consumer segments?. *European Sport Management Quarterly*, *20*(2), 130–150.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753.

Almås, I., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science*, *328*(5982), 1176–1178.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2019). Politics in the courtroom: Political ideology and jury decision making. *Journal of the European Economic Association*, *17*(3), 834–875.

Åslund, O., Hensvik, L., & Skans, O. N. (2014). Seeking similarity: How immigrants and natives manage in the labor market. *Journal of Labor Economics*, *32*(3), 405–441.

Bagues, M. F., & Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies*, *77*(4), 1301–1328.

Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2017). Does the gender composition of scientific committees matter?. *American Economic Review*, *107*(4), 1207–38.

Bar, R., & Zussman, A. (2020). Identity and bias: Insights from driving tests. *The Economic Journal*, *130*(625), 1–23.

Bennett, R., Ali-Choudhury, R., & Mousley, W. (2007). Television viewers' motivations to follow the 2005 Ashes Test series: Implications for the rebranding of English cricket. *Journal of Product & Brand Management*, *16*(1), 23–37.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review*, *96*(2), 217–221.

Bird, E. J., & Wagner, G. G. (1997). Sport as a common property resource: A solution to the dilemmas of doping. *Journal of Conflict Resolution*, *41*(6), 749–766.

Borland, J., & MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, *19*(4), 478–502.

Boyd, C. L., Epstein, L., & Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science*, *54*(2), 389–411.

Brandes, L., Franck, E., & Nüesch, S. (2008). Local heroes and superstars: An empirical analysis of star attraction in German soccer. *Journal of Sports Economics*, *9*(3), 266–286.

Brave, S. A., & Roberts, K. A. (2019). The competitive effects of performance-enhancing drugs: MLB in the posttesting era. *Journal of Sports Economics*, *20*(6), 747–781.

Buechel, B., Emrich, E., & Pohlkamp, S. (2016). Nobody's innocent: The role of customers in the doping dilemma. *Journal of Sports Economics*, *17*(8), 767–789.

Buraimo, B., Migali, G., & Simmons, R. (2016). An analysis of consumer response to corruption: Italy's Calciopoli scandal. *Oxford Bulletin of Economics and Statistics*, *78*(1), 22–41.

Camerer, C. F., & Loewenstein, G. (2004). Behavioral economics: Past, present, and future. In C. F. Camerer, G. Loewenstein, & M. Rabin (Eds.), *Advances in Behavioral Economics* (pp. 3–51). Russell Sage Foundation.

Carpenter, J. (2010). Social preferences. In S. Durlauf & L. Blume (Eds.), *Behavioural and Experimental Economics* (pp. 247–252). Palgrave Macmillan.

Chan, H. F., Savage, D., & Torgler, B. (2022). Sport as a behavioral economics lab. In H. J. R. Altman, M. Altman, & B. Torgler (Eds.), *Behavioural Sports Economics: A Research Companion* (pp. 11–51). Routledge.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.

Chavis, L., & Leslie, P. (2009). Consumer boycotts: The impact of the Iraq war on French wine sales in the US. *Quantitative Marketing and Economics*, *7*(1), 37–67.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, *99*(1), 431–57.

Chiang, Y. H., & Jane, W. J. (2013). The effects of outcome uncertainties, patriotism, and Asian regionalism in the World Baseball Classic. *Journal of Media Economics*, *26*(3), 148–161.

Cisyk, J. (2020). Impacts of performance-enhancing drug suspensions on the demand for Major League Baseball. *Journal of Sports Economics*, *21*(4), 391–419.

Cisyk, J., & Courty, P. (2017). Do fans care about compliance to doping regulations in sports? The impact of PED suspension in baseball. *Journal of Sports Economics*, *18*(4), 323–350.

Clerides, S., Davis, P., & Michis, A. (2015). National sentiment and consumer choice: The Iraq war and sales of US products in Arab countries. *The Scandinavian Journal of Economics*, *117*(3), 829–851.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281.

De Bosscher, V., Shibli, S., & De Rycke, J. (2021). The societal impact of elite sport: Positives and negatives: Introduction to ESMQ special issue. *European Sport Management Quarterly*, *21*(5), 625–635.

De Paola, M., & Scoppa, V. (2015). Gender discrimination and evaluators' gender: Evidence from Italian academia. *Economica*, *82*(325), 162–188.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter?. *American Economic Review*, *95*(2), 158–165.

Depew, B., Eren, O., & Mocan, N. (2017). Judges, juveniles, and in-group bias. *The Journal of Law and Economics*, *60*(2), 209–239.

Dimant, E., & Deutscher, C. (2019). The economics of doping in sports – A special case of corruption. In P. Downward, B. Frick, B. R. Humphreys, T. Pawlowski, J. E. Ruseski, & B. P. Soebbing (Eds.), *The SAGE Handbook of Sports Economics* (pp. 544–552). Sage.

Ding, Y., Veeman, M. M., & Adamowicz, W. L. (2013). The influence of trust on consumer behavior: An application to recurring food risks in Canada. *Journal of Economic Behavior & Organization*, *92*, 214–223.

Emerson, J. W., Seltzer, M., & Lin, D. (2009). Assessing judging bias: An example from the 2000 Olympic Games. *The American Statistician*, *63*(2), 124–131.

European Commission (2007). *White Paper on Sport* (Report No. COM(2007) 391 final).

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Feddersen, A. (2020). Do consumers care about doping scandals in sports?–Evidence from TV broadcasts of the Tour de France in Denmark. Working Paper.

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, *7*(2-3), 235–266.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The* Q*uarterly Journal of Economics*, *114*(3), 817–868.

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–Experimental evidence and new theories. In S.C. Kolm & J.M. Ythier (Eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity: Foundations, Vol. 1* (pp. 615–691). Elsevier.

Feld, J., Salamanca, N., & Hamermesh, D. S. (2016). Endophilia or exophobia: Beyond discrimination. *The Economic Journal*, *126*(594), 1503–1527.

Fershtman, C., & Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics*, *116*(1), 351–377.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369.

Frenger, M., Emrich, E., & Pitsch, W. (2013). How to produce the belief in clean sports which sells. *Performance Enhancement & Health*, *2*(4), 210–215.

Gächter, S., Herrmann, B., & Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: Survey and experimental evidence. *Journal of Economic Behavior & Organization*, *55*(4), 505–531.

Gazal-Ayal, O., & Sulitzeanu-Kenan, R. (2010). Let my people go: Ethnic in-group bias in judicial decisions—Evidence from a randomized natural experiment. *Journal of Empirical Legal Studies*, *7*(3), 403–428.

Giannetti, M., & Wang, T. Y. (2016). Corporate scandals and household stock market participation. *The Journal of Finance*, *71*(6), 2591–2636.

Giuliano, L., Levine, D. I., & Leonard, J. (2009). Manager race and the race of new hires. *Journal of Labor Economics*, *27*(4), 589–631.

Giuliano, L., Levine, D. I., & Leonard, J. (2011). Racial bias in the manager–employee relationship: An analysis of quits, dismissals, and promotions at a large retail firm. *Journal of Human Resources*, *46*(1), 26–52.

Gleaves, J., & Hunt, T. (2016). *A Global History of Doping in Sport: Drugs, Policy, and Politics*. Routledge.

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, *96*(2), 212–216.

Grossman, G., Gazal-Ayal, O., Pimentel, S. D., & Weinstein, J. M. (2016). Descriptive representation and judicial outcomes in multiethnic societies. *American Journal of Political Science*, *60*(1), 44–69.

Gruhl, J., Spohn, C., & Welch, S. (1981). Women as policymakers: The case of trial judges. *American Journal of Political Science*, *25*(2), 308–322.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance*, *63*(6), 2557–2600.

Guiso, L., Sapienza, P., & Zingales, L. (2009). Cultural biases in economic exchange?. *The Quarterly Journal of Economics*, *124*(3), 1095–1131.

Güth, W., Ockenfels, P., & Wendel, M. (1993). Efficiency by trust in fairness? Multiperiod ultimatum bargaining experiments with an increasing cake. *International Journal of Game Theory*, *22*(1), 51–73.

Güth, W., Ockenfels, P., & Wendel, M. (1997). Cooperation based on trust: An experimental investigation. *Journal of Economic Psychology*, *18*(1), 15–43.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388.

Haugen, K. K. (2004). The performance-enhancing drug game. *Journal of Sports Economics*, *5*(1), 67–86.

Hausman, J. A., & Leonard, G. K. (1997). Superstars in the National Basketball Association: Economic value and policy. *Journal of Labor Economics*, *15*(4), 586–624.

Heilmann, K. (2016). Does political conflict hurt trade? Evidence from consumer boycotts. *Journal of International Economics*, *99*, 179–191.

H.M. Government (Her Majesty's Government) (2015). *Sporting Future: A New Strategy for an Active Nation*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/486622/Sporting_Future_ACCESSIBLE.pdf

Hoekstra, M., & Street, B. (2021). The effect of own-gender jurors on conviction rates. *The Journal of Law and Economics*, *64*(3), 513–537.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, *76*(4), 728–741.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Kamas, L., & Preston, A. (2012). Distributive and reciprocal fairness: What can we learn from the heterogeneity of social preferences?. *Journal of Economic Psychology*, *33*(3), 538–553.

Klein, J. G., Ettenson, R., & Morris, M. D. (1998). The animosity model of foreign product purchase: An empirical test in the People's Republic of China. *Journal of Marketing*, *62*(1), 89–100.

Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics*, *112*(4), 1251–1288.

Knepper, M. (2018). When the shadow is the substance: Judge gender and the outcomes of workplace sex discrimination cases. *Journal of Labor Economics*, *36*(3), 623–664.

Konjer, M., Meier, H. E., & Wedeking, K. (2017). Consumer demand for telecasts of tennis matches in Germany. *Journal of Sports Economics*, *18*(4), 351–375.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature*, *41*(4), 1188–1239.

Laverie, D. A., & Arnett, D. B. (2000). Factors affecting fan attendance: The influence of identity salience and satisfaction. *Journal of Leisure Research*, *32*(2), 225–246.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*(3), 593–622.

Lim, C. S., Silveira, B. S., & Snyder, J. M. (2016). Do judges' characteristics matter? Ethnicity, gender, and partisanship in Texas state trial courts. *American Law and Economics Review*, *18*(2), 302–357.

Loland, S (2001). *Fair Play in Sport: A Moral Norm System*. Routledge.

Lyngstad, T. H., Härkönen, J., & Rønneberg, L. T. S. (2020). Nationalistic bias in sport performance evaluations: An example from the ski jumping world cup. *European Journal for Sport and Society*, *17*(3), 250–264.

MacDonald, G. M. (1988). The economics of rising stars. *American Economic Review*, *78*(1), 155–166.

Manoli, A. E., Bandura, C., & Downward, P. (2020). Perceptions of integrity in sport: Insights into people's relationship with sport. *International Journal of Sport Policy and Politics*, *12*(2), 207–220.

Mayda, A. M., & Rodrik, D. (2005). Why are some people (and countries) more protectionist than others?. *European Economic Review*, *49*(6), 1393–1430.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, *23*(3), 473–490.

Mongeon, K., & Longley, N. (2015). Testing for ethnicity discrimination among NHL referees: A duration model approach. *Eastern Economic Journal*, *41*(1), 86–101.

Nalbantis, G., & Pawlowski, T. (2019). US Demand for European soccer telecasts: a between-country test of the uncertainty of outcome hypothesis. *Journal of Sports Economics*, *20*(6), 797-818.

Neale, W. C. (1964). The peculiar economics of professional sports. *The Quarterly Journal of Economics*, *78*(1), 1–14.

Nüesch, S., & Franck, E. (2009). The role of patriotism in explaining the TV audience of national team games–Evidence from four international tournaments. *Journal of Media Economics*, *22*(1), 6–19.

Page, L., & Page, K. (2010). Evidence of referees' national favouritism in rugby. NCER Working Paper Series 62, NCER, 1–35.

Pandya, S. S., & Venkatesan, R. (2016). French roast: Consumer response to international conflict—Evidence from supermarket scanner data. *Review of Economics and Statistics*, *98*(1), 42–56.

Parsons, C. A., Sulaeman, J., Yates, M. C., & Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review*, *101*(4), 1410–35.

Pawlowski, T., & Nalbantis, G. (2019). Competitive balance: Measurement and relevance. In P. Downward, B. Frick, B. R. Humphreys, T. Pawlowski, J. E. Ruseski, & B. P. Soebbing (Eds.), *The SAGE Handbook of Sports Economics* (pp. 154–162). Sage.

Pawlowski, T., Nalbantis, G., & Coates, D. (2018). Perceived game uncertainty, suspense and the demand for sport. *Economic Inquiry*, *56*(1), 173–192.

Pope, B. R., & Pope, N. G. (2015). Own-nationality bias: Evidence from UEFA Champions League football referees. *Economic Inquiry*, *53*(2), 1292–1304.

Pope, D. G., Price, J., & Wolfers, J. (2018). Awareness reduces racial bias. *Management Science*, *64*(11), 4988–4995.

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

Putnam, R. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*(5), 1281–1302.

Riefler, P., & Diamantopoulos, A. (2007). Consumer animosity: A literature review and a reconsideration of its measurement. *International Marketing Review*, *24*(1), 87–119.

Rodríguez, C., Pérez, L., Puente, V., & Rodríguez, P. (2015). The determinants of television audience for professional cycling: The case of Spain. *Journal of Sports Economics*, *16*(1), 26–58.

Rosen, S. (1981). The economics of superstars. *American Economic Review*, *71*(5), 845–858.

Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy*, *64*(3), 242–258.

Sandberg, A. (2018). Competing identities: A field study of in-group bias among professional evaluators. *The 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), 708–734.

Schreyer, D., Schmidt, S. L., & Torgler, B. (2018). Game outcome uncertainty in the English Premier League: Do German fans care? *Journal of Sports Economics*, *19*(5), 625–644.

Schreyer, D., & Torgler, B. (2018). On the role of race outcome uncertainty in the TV demand for Formula 1 Grands Prix. *Journal of Sports Economics*, *19*(2), 211–229.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review*, *103*(2), 147–174.

Shayo, M. (2020). Social identity and economic policy. *Annual Review of Economics*, *12*, 355–389.

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

Shtudiner, Z., Tohar, G., & Kantor, J. (2021). The effect of identification with a sports team and its performance on the willingness of fans to pay for team products. *Managerial and Decision Economics*, *43*(3), 607–615.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Wadsworth.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7–24). Nelson-Hall.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

UNESCO (2005). *International Convention Against Doping in Sport* (Report No. ED.2005/CONVENTION ANTI-DOPING Rev).

Van Reeth, D. (2013). Television demand for the Tour de France: The importance of outcome uncertainty, patriotism and doping. *International Journal of Sport Finance*, *8*(1), 39–60.

Verlegh, P. W., & Steenkamp, J. B. E. (1999). A review and meta-analysis of country-of-origin research. *Journal of Economic Psychology*, *20*(5), 521–546.

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

Zitzewitz, E. (2014). Does transparency reduce favoritism and corruption? Evidence from the reform of figure skating judging. *Journal of Sports Economics*, *15*(1), 3–30.