Out of Sight, Out of Mind? The Effect of Peer Evaluators on In-Group Bias*

A. Nilesh Fernando & Siddharth Eapen George †

September 4, 2022

Abstract

We study how peers affect in-group bias. Exploiting several umpiring reforms in international cricket matches —where two umpires make independent decisions in each other's presence—we show that home-team umpires are less biased when working with a neutral colleague, i.e. one who is neither a national of the home nor the foreign team. This temporary debiasing is driven by the social pressure umpires feel to be impartial in the presence of neutral peers. Performance evaluation by visually non-salient monitors does not reduce bias, suggesting that physical presence is an important component of debiasing and peer influence.

JEL Classification Numbers: J71, J15, D91

Keywords: Discrimination, In-group bias, Peer influence, Social Pressure, Cricket

*This paper was previously circulated as "Debiasing Discriminators: Evidence from the Introduction of Neutral Umpires"

†University of Notre Dame, Notre Dame, IN (nilesh.fernando@nd.edu) and National University of Singapore (segeorge@nus.edu.sg). The data used in this paper was obtained from publicly available statistics on ESPNCricinfo.com and the processed datasets are available upon request. This research received financial support from the Institute for Scholarship in the Liberal Arts (ISLA) at the University of Notre Dame. George is the co-founder of a sports technology start-up (SeeHow Pte Ltd) that has developed a smart cricket ball. We thank Chungeun Yoon, Patrick Hollander and Pranjal Bhushan for research assistance. We are grateful to Taryn Dinkelman, William Evans, Lakshmi Iyer, Lawrence Katz, Asim Khwaja, Matt Notowidigdo, Sandip Sukhtankar, Amanda Pallais, Rohini Pande, and Abigail Wozniak for helpful suggestions. We also thank seminar participants at Harvard, Michigan, Notre Dame, and Middlebury. All errors remaining are our own.

1 Introduction

Many organizations task committees of peers with making impartial decisions. Corporate boards determine hiring and compensation policies in firms. Faculty committees make hiring and tenure decisions in universities. Judicial panels deliver verdicts. Though directed to act with impartiality, an extensive literature suggests that committees are prone to in-group biases.[1] Concerns about bias have led many organizations to promote diversity among influential committees. Yet a recent review of the literature concludes that there is relatively little work examining how diversity initiatives influence bias and it is especially difficult to establish the underlying mechanisms (Bertrand and Duflo, 2017).

In this paper, we study how peers affect in-group bias among professional umpires in international cricket matches. Cricket provides an excellent lab to investigate this question, owing to a series of reforms that mandated diversity on umpiring panels. In addition, cricket umpires make independent decisions with varying amounts of discretion, allowing us to isolate changes driven by umpire responses using detailed decision-level data. We primarily examine the impacts of a reform that mandated the presence of a *neutral umpire* in international cricket matches. Prior to the reform all matches were refereed by two umpires, both of whom were from the host country (hereafter, *home umpires*). In 1994, a reform mandated that a neutral umpire —who shares the nationality of neither the home team nor the foreign team—be present in certain types of matches. Our main finding is that home umpires temper their in-group bias when they are in the *physical presence* of a neutral umpire.

While the neutral umpire reform affected Test matches, it did not apply to One Day Internationals (ODIs), the other main class of matches, which continued to be solely refereed by home umpires. This variation motivates a difference-in-difference design which compares matches *before* versus *after* the reform in *Test* versus *ODI* matches. Our primary measure of bias is the number of high-stakes discretionary decisions that umpires made against the home versus against the foreign team.[2] Compliance with the reform is almost perfect and there are no pre-trends, allowing us to cleanly estimate the causal effect of a neutral umpire.

We find that the presence of a neutral umpire significantly reduced bias towards the home team. Panels with a neutral umpire award 14% *more* discretionary decisions against the home team and 23% *fewer* discretionary decisions against the foreign team. The reform did not affect decisions where umpires have little discretion, suggesting that the treatment effects are driven by changes in umpire behavior rather than endogenous player responses to the policy change. The reduction in bias could be driven by (i) a *replacement effect*, where impartial neutral umpires replace

[1] In-group biases has been show in contexts as varied as juries (Anwar, Bayer, and Hjalmarsson, 2012), hiring committees (Bagues and Esteve-Volart, 2010), and professional sports (Price and Wolfers, 2010a).

[2] We focus on decisions where umpires judge whether a batsman is out and describe our classification of decisions in more detail in section 2.3.1

biased home umpires, (ii) a *debiasing effect*, where home umpires are less biased in the presence of a neutral umpire, or both.

To distinguish between these explanations, we use decision-level data on a sample of matches where we can identify the umpire responsible for each decision. Home umpires generally display significant in-group bias, awarding 14% more decisions against foreign teams. However, when paired with a neutral umpire, this gap is virtually eliminated: in the presence of a neutral colleague, home umpires award 12% fewer decisions against foreign teams and we cannot reject that they are unbiased. These findings suggest that neutral colleagues have a strong *debiasing* impact on home umpires.

Why might an evaluator be less biased in the presence of a neutral peer? Each umpire makes individual decisions (so there are no incentives to coordinate or achieve unanimity) and decisions are made independently with no input from each other (so there are no deliberation effects). The absence of these channels—in addition to evidence suggesting no changes in preferences or beliefs—suggests that the reform may have influenced how umpires behave simply because they are in the presence of a neutral colleague. Following a literature in psychology (Guerin, 1986), we refer to these as *presence effects*, since they capture the impact of social pressure or image concerns due to a peer being physically present.

We provide four pieces of evidence which collectively suggest that the physical presence of a neutral colleague exerts social pressure on home umpires to be impartial. First, we show that home umpires are less susceptible to crowd pressure when working with a neutral umpire. Home umpires give 14% more decisions against foreign teams during weekend days of the same match (when crowds are larger), indicating a general vulnerability to crowd pressure, consistent with results in other contexts.[3] However, the presence of a neutral colleague largely eliminates the influence of crowd pressure on home umpires.

Second, we find that the debiasing effects of neutral umpires are temporary. Home umpires who work with a neutral colleague are significantly less biased during that match, but their behavior in subsequent matches, even those held just a few days later, is unaffected. By contrast, most studies of the contact hypothesis find that interactions with other groups lead to persistent reductions in bias (Bazzi, Gaduh, Rothenberg, and Wong, 2019; Lowe, 2021). This channel is thus unlikely to be driving our results.

Next, exploring heterogeneity, we find that neutral umpires are more effective at curtailing bias when home umpires have strong reputational reasons to appear impartial. Neutral umpires have greater impacts when (i) the home umpire has a track record of bias before the reform, (ii) the neutral umpire is an experienced senior official and (iii) both umpires have not previously worked together. These patterns suggest that home umpires curb their bias when they value their

[3] Garicano, Palacios-Huerta, and Prendergast (2005), Endrich and Gesche (2020), Bryson, Dolton, Reade, Schreyer, and Singleton (2021) and Cohen, Neeman, and Auferoth (2021) document how crowd pressure causes referees to favour home teams in soccer matches.

colleagues' approval and are keen to present a good image.

Finally, we study how external monitors influence umpire bias by examining the impact of a separate reform that introduced neutral monitors (called match referees) to evaluate umpires' performance and report unethical behavior to the International Cricket Council. These performance reports were directly linked to umpires' promotion and contract renewal. We find that this reform had a limited (detectable) impact on bias suggesting that career concerns *alone* are unlikely to be driving the debiasing effect of neutral peers.

Taken together, these results suggest that the physical presence of a neutral colleague exerts social pressure on individuals to temporarily moderate their biases. The separate introduction of neutral *on-field* umpires and *off-field* match referees enable us to assess the relative importance of physical presence in driving debiasing. Our evidence points to the significant role that *presence effects* may play in mitigating in-group bias and in peer influence more generally.

Presence effects imply that small changes to peer composition can have large effects on bias. One implication of this proposition for anti-bias reform is that it may not be necessary to remove all conflicted evaluators in order to eliminate bias. To test this implication, we analyze a subsequent reform that required both umpires in Test matches to be neutral (ie. going from one to two neutral umpires), and find that it did not reduce bias further. In support of the external validity of this finding, we re-analyze data from Neggers (2018), who studies in-group bias among election officials in India. The paper's key result is that parties representing religious and ethnic minorities receive more votes at polling booths with minority election officials. Re-analyzing his data, we find that this effect is entirely driven by the presence of one minority election official. As in our setting, a second minority official has no additional impact. This pattern is consistent with majority-group officials curbing their in-group bias when in the presence of a minority-group colleague.

We primarily contribute to a literature examining how peers influence discriminatory behavior. Several papers have examined how the composition of important committees like juries, judicial panels and hiring committees affect racial, gender and in-group bias.[4] In the world of professional sports, Price and Wolfers (2010b) show that white players receive fewer fouls in matches with more white referees, but cannot distinguish compositional changes from peer influence. Sandberg (2018) shows that judges in international dressage are biased towards co-nationals of peer judges. This bias transmission effect, which is contrary to our findings, may be driven by rules encouraging consensus among dressage judges that are not present in cricket. Overall, these papers offer mixed conclusions, with diverse committees reducing in-group bias in some cases (Anwar, Bayer, and Hjalmarsson, 2012) while intensifying bias in others (Bagues, Sylos-Labini, and Zinovyeva, 2017). The underlying mechanisms are also unclear, since these committees can deliberate and often have incentives to reach unanimous decisions. Our findings, instead, suggest

[4]See, for example, (Anwar, Bayer, and Hjalmarsson, 2012; Bagues and Esteve-Volart, 2010; De Paola and Scoppa, 2015; Grossman, Gazal-Ayal, Pimentel, and Weinstein, 2016; Flanagan, 2018; Hoekstra and Street, 2021).

that presence effects play an important role in reducing bias. This suggests a different channel through which inter-group contact may reduce bias, albeit temporarily, even if underlying beliefs and preferences are unchanged.

Second, our work contributes to a growing literature on social pressure. Social image concerns shape behavior in many organizations (Mas and Moretti, 2009; Bursztyn and Jensen, 2017) and may also affect the public expression of bias (Bursztyn, Egorov, and Fiorin, 2020). Prior work has shown that in-group bias, while a pervasive feature of human behavior (Shayo, 2020), is seen as desirable in some social environments but stigmatized in others (Barr, Lane, and Nosenzo, 2018; Harris, Herrmann, Kontoleon, and Newton, 2015). We make two contributions to this literature. First, we show that social pressure has dynamic effects that vary with relationship length—in our setting, debiasing is strongest when umpires have not worked together before. Second, we provide evidence that social pressure has weaker effects when individuals face identity or self-image costs for acquiescing to social pressure. Being labeled a *neutral* umpire may increase the salience of an individual's professional identity, raise compliance with professional norms like impartiality and reduce influence from crowds and peers.

Finally, our paper relates to a literature that uses sports to study economic behavior. Guryan, Kroft, and Notowidigdo (2009) estimates peer effects among professional golfers, exploiting features of the sport to overcome identification challenges such as correlated shocks. List (2004) exploits features of the sportscard market to understand the underlying nature of bias and distinguish statistical from taste-based discrimination. Similarly, features of cricket enable us to identify presence effects and rule out alternative mechanisms through which peers may influence bias. Many studies show that referees display bias along multiple dimensions ((Dohmen and Sauermann, 2016; Garicano, Palacios-Huerta, and Prendergast, 2005; Feess, Mueller, et al., 2021; Cohen, Neeman, and Auferoth, 2021). Sacheti, Gregory-Smith, and Paton (2015) is the only other paper studying umpiring bias in cricket: they describe how home bias has declined over time and attribute this to the 1994 neutral umpire reform. However, their empirical approach conflates all time-varying factors (including contemporaneous reforms) that affect the performance of foreign teams, making it difficult to isolate the causal effect of any specific reform. In addition, they do not study peer influence, whereas our core finding is the debiasing effect of neutral peers. We also exploit other natural experiments, like the introduction of off-field monitors, to show that our findings are best explained by presence effects and consider the external validity of our findings outside sports.

The rest of the paper is organized as follows. The next section discusses our contributions to the literature. Section 2 explains the great game of cricket, our bias measures and the umpiring reforms that we rely on for identification. Section 3 describes the data and empirical strategy. Section 4 reports the baseline results, and Section 5 discusses mechanisms. Section 6 examines the robustness of our main results, and Section 7 concludes.

2 Institutional Background

2.1 Basics of Cricket

Cricket is a bat-and-ball sport that originated in south-east England in the 16th century. The sport diffused across the globe as the British empire expanded, and it is now the world's second most popular sport, played by 106 countries and followed by 3 billion fans. The game is played between two teams of 11 players. Each team takes a turn to bat—an *innings*—while the other team bowls. The batting team tries to score as many *runs*—i.e. points—as possible, while the bowling team attempts to get batsmen out and restrict run-scoring. An innings ends when the stipulated number of balls has been bowled[5] or when all batsmen are out. The team that scores more runs wins.[6]

2.2 International Cricket Matches

While domestic leagues and franchise tournaments are increasingly popular, international cricket matches between the national teams of different countries remain the most prestigious and frequent type of competition. The first international matches were held in the late 1800s between England and Australia. Over 100 countries are members of the International Cricket Council, the sport's governing body, though only the 12 full members play Test matches, the oldest and longest form of the game.

Most matches occur during bilateral series between two countries with one nation hosting (the *home team*) and another nation visiting (the *foreign team*). As in other settings, national identity is the most salient group identity during these contests (Depetris-Chauvin, Durante, and Campante, 2020), and the dimension along which we study in-group bias.

There are 3 types of international matches—Test matches, One Day Internationals (ODIs) and Twenty20s (T20s). All categories follow exactly the same rules of play and differ primarily in their length. Test matches last 5 days; during this period, each team has 2 innings to bat, with an unlimited number of overs for each inning. By contrast, ODIs and T20s are *limited overs* matches—each team has only 1 batting inning, which lasts 50 overs in ODIs and 20 overs in T20s. Our sample restricts attention to Tests and ODIs as T20s began after the umpiring reforms we study.

2.3 Umpiring in Cricket

All international matches are refereed by two on-field umpires. Both umpires are present throughout the match, but only one umpire is primarily responsible for decision-making at any time. Um-

[5]The length of a cricket match is measured in *overs*, which consist of 6 legitimate balls from the bowler.

[6]The batting team is represented on the field by two players (*batsmen*) at a time. One batsman (the *striker*) stands at one end of the 22-yard pitch facing the bowler and tries to hit the ball. The other batsman (the *non-striker*) stands at the other end of the pitch. Runs can be scored in two ways: (i) when the batsman hits the ball across the boundary, or (ii) when the two batsman cross and swap ends of the pitch. Each time the batsmen run and swap ends counts as one run.

pires alternate charge every over (6 balls), with odd overs being refereed by one umpire and even overs by the other umpire. The umpire in charge of an over stands at one end of the pitch next to the bowler, giving him a clear view of the wickets, the batsman and the ball's trajectory.

The second umpire stands about 30 yards away, perpendicular to the pitch, at a position on the field called *square leg*. While in the direct eye-line of the primary umpire, square leg does not offer a good angle to judge the quality of the primary umpire's decisions.[7] At the end of the over, the umpires switch charge and positions, and alternate in this manner for the rest of the innings.

The umpires almost never deliberate on decisions, though they have informal social contact before and after the match and (especially during 5-day Test matches) over lunch and tea breaks.

Overall, we believe these features provide an excellent, real-world laboratory to study presence effects: professional evaluators make thousands of independent high-stakes decisions in the presence of a colleague who cannot actually assess the quality of their decisions. We study how the colleague's identity affects bias.

2.3.1 Measuring Bias: Discretionary vs Non-discretionary decisions

Our measure of bias is the number of high-stakes decisions an umpire gives against the home versus against the foreign team. The most significant call a cricket umpire makes is deciding whether to give a batsman out.[8] Batsmen can get out in several ways, and some dismissal categories offer umpires significantly more discretion than others. Appendix A1 provides descriptions of non-discretionary and discretionary decisions, and Appendix A2 presents visual illustrations of some common decisions.

Identifying discretionary and non-discretionary decisions offers two advantages. First, we can study how umpiring reforms affect discretionary decisions, which are more directly affected by umpiring bias. Second, non-discretionary decisions enable us to construct *placebo outcomes* that capture other channels through which the reforms might affect cricket matches (e.g. player responses). Showing that the reform only affected discretionary decisions supports our interpretation that the treatment effects are driven by changes in umpire behavior.

2.4 Umpiring Reforms

Historically, all international matches were refereed by two umpires from the home country. This system persisted despite obvious concerns about bias. In 1986, Pakistan's cricket captain, Imran Khan, irritated that his team's strong home record was attributed to biased umpiring, insisted that two Indian umpires officiate a match in Lahore against the dominant West Indies side. He repeated the gesture in a high-profile series against India in 1989, inviting two English umpires to

[7] The square leg umpire has a clear view of the crease at the batsman's end, and therefore sometimes makes run out and stumping decisions.

[8] A batsman has only one chance to bat in an inning and each team has only 1 or 2 innings, depending on the match type, so giving a batsman out is a key decision and the main outcome we study.

referee.

Mandating One Neutral Umpire: In 1994, the ICC institutionalized this idea, mandating that Test matches were to be umpired by one home umpire and one *neutral umpire*—i.e. an umpire who was a national of neither playing team. ODIs were unaffected, and continued to be refereed by two home umpires.

Second Neutral Umpire: In 2002, the ICC extended this rule, requiring both on-field umpires in Test matches to be neutral. Panel A of Figure 1 illustrates the near-perfect compliance with both the 1994 and 2002 reforms.

Match Referees and TV umpires: In 1992, the ICC introduced TV umpires and match referees to support and monitor the on-field umpires. Panel B of Figure 1 shows that match referees were adopted rapidly in both Tests and ODIs over a short period. TV umpires have real-time access to slow-motion replays, and the on-field umpires could refer close-call run out and stumping decisions to the TV umpire to adjudicate on.

The match referee has a broader role: he oversees the overall governance of the match, ensures that the ICC Cricket Code of Conduct is followed, and hands out penalties for any breaches. The match referee has no direct involvement in the match and does not make any decisions that influence the course of play. However, after each game, the match referee submits a report to the ICC which notes any improper behavior by players or officials and awards a numerical performance score to each on-field umpire. These scores are used by the ICC to appraise umpires. Strong performers can be promoted to list of elite umpires, while weak performers may have their contracts terminated. As such, we consider the introduction of match referees akin to a monitoring intervention that increased the career concerns of the on-field umpires.

3 Data and Empirical Strategy

3.1 Data

3.1.1 Match-level Data

We scraped data on international cricket matches from ESPNCricinfo.com, a reputed cricket news and statistics website. Our data covers the universe of international matches (N=5020) over the period 1877-2011.[9] For each match, we have information on all umpiring decisions (both discretionary and non-discretionary), other match outcomes (e.g. runs scored) and the identity and

[9]The sample includes 1,986 Test Matches and 3,034 ODIs. Appendix A3 charts the number of matches over time by match type.

nationality of the umpires and match referee. Because our policy variation— the presence of neutral umpires —is at the match level, we aggregate our data at the team-match level in our baseline analysis, yielding 10,040 observations.[10] This allows us to separately analyze how the reform affected decisions against home and foreign teams.

3.1.2 Decision-level Data

The match-level data allows us to estimate the overall impact of the neutral umpire policy on decisions against the home and foreign teams. However, in order to test for peer influence (i.e. whether home umpires are less biased when working with neutral umpires), we need to observe individual decisions. Indeed, the lack of referee-level data prevents other papers (e.g. Price and Wolfers (2010b); Sacheti, Gregory-Smith, and Paton (2015)) from estimating peer influence.

We compile decision-level data for a subset of matches where we are able to identify the umpire responsible for each decision and, in particular, whether he was a home or neutral umpire. We scraped ball-by-ball commentary records and constructed an algorithm to analyze the commentary text data to identify which umpire was responsible for each decision.[11] We focus on LBW dismissals in the decision-level analysis, as run out and stumping decisions are sometimes made by the square leg umpire.

Overall, we are able to identify the affiliation of the decision umpire for 109,996 decisions in 3891 matches. We collapse the decision-level data at the umpire-innings level, and study how the behavior of home and neutral umpires is influenced by the identity of their peer umpire.

3.1.3 Summary Statistics

Table 1 presents summary statistics. In Panel A, data is collapsed at the match-team level and shows the average number of decisions each team experiences in a match. The most common type of dismissal is caught in the outfield, with approximately 5.25 such dismissals per match for each team. On average, a team receives about 3 discretionary dismissals per match and 7.5 non-discretionary dismissals.

In Panels B and C, we collapse data at the match level, and present summary statistics on match and umpire characteristics. Our sample contains data on 5020 international matches, 40% of which are Test matches and 60% are ODIs. There appears to be significant home advantage, with foreign teams winning only 13% of the time.[12]

Panel C presents summary statistics on umpire characteristics. 44% of matches are officiated

[10] Each observation is a match-level outcome for either the home or foreign team.

[11] Specifically,our algorithm identifies which umpire officiated even overs and which umpire was in charge during odd overs. We then attribute dismissals that occurred in even and odd overs to the relevant umpire, since umpires alternate charge ever over throughout an innings. We validate the robustness of our algorithm by manually coding decisions for a sample of matches.

[12] Home teams win 47% of the time, with drawn matches comprising the remainder.

by two home umpires; 1 neutral umpire is present in 25% of matches and 30% of matches are refereed by 2 neutral umpires.[13] As we will show, this variation is mostly driven by the umpiring reforms we described in the previous section. A match referee and TV umpire are present in a majority (57%) of matches and in the vast majority of matches after 1992.

3.2 Empirical strategy: Match-Level Effects of the Reform

We first examine how the presence of neutral umpires affects in-group bias by analyzing the umpiring reforms discussed in Section 2.4. Our analysis begins by focusing on the 1994 reform that mandated the presence of one neutral umpire in Test matches but did not affect ODI matches. Our empirical strategy is a difference-in-differences (DiD) design that compares (i) *Test* vs *ODI* matches, (ii) *before* vs *after* the reform. We separately examine how the reform influenced decisions against home and foreign teams. Specifically, we estimate the following regression equation:

$$Y_{impgt} = \beta_1 Test_m + \beta_2 PostReform_t + \beta_3 (Test\ x\ PostReform)_{mt} + \alpha_{pt} + \alpha_g + Controls_{mt} + \varepsilon_{impgt} \quad (1)$$

where Y_{impgt} denotes an outcome (e.g. discretionary decisions) for team i in match m between teams p on ground g in year t. $Test_m$ is a dummy for Test matches (the *treated* group), and $Post_t$ is a dummy for matches held after the umpiring reform. β_3, the DiD estimand, is the coefficient of interest. Our baseline specification includes two types of fixed effects, though our results are robust to excluding these or including other fixed effects. α_{pt} refers to *Team pair x Year FEs*, which control for the quality of the playing teams in each year. α_g are *ground FEs*, which capture the impact of unobserved factors associated with a particular stadium or location (e.g. climatic characteristics) that influence match outcomes. We include several match-level controls, such as whether the match has a TV umpire and match referee, and whether the home team won the toss.[14] Standard errors are clustered at the match level.

3.2.1 Compliance with Reform and Pre-Trends

We provide two pieces of evidence to validate our DiD design. First, we demonstrate high compliance with the umpiring reforms. Second, we provide evidence to support the parallel trends assumption.

Panel A in Figure 1 shows the average number of neutral umpires officiating matches over time. The blue line plots the number of neutral umpires in Tests while the green line plots the

[13] A very small sample of matches (about 100 matches or 1% of the whole sample) are refereed by 1 home and 1 away umpire, i.e. an umpire from each playing team.

[14] The team that wins the toss decides whether to bat or bowl first. Teams usually exploit this choice to make strategic use of favorable playing conditions. For example, when the sky is overcast, the ball typically swings more in the air, making it more difficult to bat. Winning the toss thus confers an advantage. Indeed, we find that the home team is 6pp more likely to win (the baseline rate is 47%) when they win the toss.

figure for ODIs. Prior to 1994, over 96% of matches had no neutral umpires. In 1994, the year of the first reform, we observe a sudden increase for Test matches: now over 91% of Test matches have 1 neutral umpire, but we see no change for ODI matches. After the sharp increase in 1994, we see no discernible changes for either Test or ODI matches until 2002, when additional reforms came into effect. Post 2002, we observe that all Test matches have 2 neutral umpires and ODIs have 1 neutral umpire, again consistent with the relevant reforms. Finally, we see that match referees and TV umpires are adopted very rapidly after their introduction in 1992. On the whole, this indicates very strong compliance with the umpiring reforms.

Our DiD design identifies the causal effect of the neutral umpire policy under the (standard) assumption that outcomes for Test and ODI matches had parallel trends before the reform. To validate the parallel trends assumption, we plot the year-by-year coefficients from equation (1), comparing discretionary decisions in Tests versus ODIs. We plot these coefficients in Appendix B2. Figures B2a and B2b show the coefficients for the home and foreign team respectively.[15] We see that the lags are close to zero in Panel A, while after the reform we find a relative increase in discretionary decisions for the home team in Test matches. We cannot reject that all year coefficients are equal to zero prior to the 1994 reform: for our primary outcome, the inverse hyperbolic sine (IHS) transformation of the number of discretionary decisions, the joint F-test is 0.93 (p-value=0.52) for home team outcomes and 1.01 (p-value=0.44) for foreign team outcomes.[16] Thus, both visual inspection of the event study graphs and econometric testing of the pre-trend coefficients both support the validity of the parallel trends assumption in our setting.

3.3 Empirical strategy: Effects of the Reform by Umpire Type

We collapse the decision-level data described in Section 3.1.2 to the umpire-inning level to identify the umpire-level effects of the reform. In particular, using variation in the composition of umpiring panels due to both the 1994 reform and the 2002 reform (two neutral umpires, see section 2.4, we examine whether home umpires give fewer discretionary decisions against foreign teams when paired with a neutral umpire rather than a fellow home umpire. We estimate these effects using the following regression:

$$Y_{ijmpgt} = \beta_1 Foreign_i + \beta_2 NeutralUmpPresent_m + \beta_3 (Foreign_i \; x \; NeutralUmpPresent_m) + \alpha_{pt} + \alpha_g + \alpha_j + X_m + \varepsilon_{impgt} \quad (2)$$

where Y_{impgt} denotes LBW dismissals given against team i by umpire j in match m that is played between teams p on ground g in year t. $Foreign_i$ is a dummy for decisions against the

[15] In both Panel A and B, the excluded year is 1993 and the specification includes team-pair FEs, year FEs, ground FE, and a control for the presence of a television umpire and match referee as in equation (1).

[16] Using the raw number of discretionary decisions as the outcome variable, the joint F-test is 1.15 (p-value = 0.31) for home team outcomes and 1.27 (p-value = 0.23) for foreign team outcomes.

foreign team, $NeutralUmpPresent_m$ is a dummy for whether the decision umpire j is paired with a neutral umpire, and β_3, the coefficient on their interaction, is the main coefficient of interest. As in the the baseline difference-in-difference specification, we include team pair x year FEs (α_{pt}), ground FEs (α_g) and match controls (X_m), but also add umpire FEs (α_j) to control for unobserved umpire characteristics that may affect bias or the frequency of LBW decisions.

4 Results

4.1 The Impact of Introducing Neutral Umpires

We first present results on the impact of the 1994 reform, which required one neutral umpire in Test matches but did not affect ODI matches. Table 2 presents results from our baseline DiD specification. We see that the introduction of a neutral umpire *increased* discretionary decisions against the home team by 14% (column 1, significant at the 5% level), while *reducing* discretionary decisions against foreign teams by 23% (column 2, significant at the 1% level). Non-discretionary decisions were unaffected, with point estimates for both home and foreign teams less than 1% in magnitude and statistically insignificant (columns 3-4). Appendix table B3 shows how the reform impacted each type of discretionary and non-discretionary decision. We see a clear pattern where the introduction of a neutral umpire leads to more discretionary decisions against home teams (average effect of +11%), fewer discretionary decisions against foreign teams (average effect of -10%), and has no impact on non-discretionary decisions for either team (average effect of 1.6%). The fact that we only observe changes for discretionary decisions suggests that the reform changed umpire rather than player behavior.

The reform also affected overall match outcomes. Matches became more competitive: 8pp more matches are close, a sizeable increase from the pre-reform baseline of 30%.[17] In addition, foreign team were 6pp more likely to win, a non-trivial increase given the baseline win probability of 13%, but this effect is not precisely estimated (p=0.17).[18] The effect is driven by matches in which the foreign team was lucky and won the toss: in such matches, the reform raised foreign teams' win rates by 17pp (from 30% to 47%), nearly eliminating home advantage.

These results suggest that the introduction a neutral umpire significantly reduced bias toward the home team. Next, we explore why the introduction of neutral umpires reduced bias.

4.2 The Debiasing Effect of Neutral Peers on Home Umpires

Our match-level results could either be driven by (i) a *replacement effect*, where unbiased neutral umpires replace partisan home umpires and/or (ii) a *debiasing effect*, where home umpires are less

[17] A close game is defined as a match with a below median pre-reform victory margin in terms of number of runs or wickets.

[18] We attribute this imprecision to the fact that discretionary decisions typically only account for about 30% of all dismissals.

biased when working with a neutral umpire colleague. It is not possible to distinguish these two channels using match-level data. To do so, we turn to the decision-level data, which allows us to identify the decision umpire for all LBW decisions in a subset of matches.[19]

4.2.1 Descriptive Patterns from Decision-Level Data

Appendix table B1 presents descriptive patterns from the raw data. We see that home umpires, when paired with another home umpire, give more LBW decisions against foreign teams than home teams (0.65 vs 0.47 decisions per match), with the difference significant at the 1% level. However, when paired with a neutral umpire, home umpires no longer give more decisions against the foreign team. By contrast, neutral umpires award a similar number of LBW decisions against home and foreign teams, and this behavior is unaffected by whether the neutral umpire is paired with a home umpire or another neutral umpire. We now investigate these suggestive patterns more formally.

4.2.2 Are Home Umpires Less Biased when Paired with a Neutral Umpire?

Table 3 presents results from estimating equation (2). Columns 1 and 2 focus on decisions made by home umpires. At baseline, we see that home umpires display significant in-group bias, awarding 14% more LBW decisions per match against foreign teams. However, this difference is virtually eliminated when home umpires are paired with a neutral umpire: the presence of a neutral colleague causes a home umpire to award 13% fewer decisions against the foreign team. This result is robust to the inclusion of umpire fixed effects (column 2), indicating that unobserved changes in umpire characteristics cannot explain this effect. Rather, the same home umpire is less biased when working with a neutral peer.

Columns 3 and 4 focus on decisions by neutral umpires. First, we notice that neutral umpires on average do not display bias toward home teams: they award only 4-5% more decisions against foreign teams and this difference is not statistically significant. Moreover, neutral umpires' behavior is not affected by whether their on-field colleague is a home or neutral umpire: the coefficient on *Foreign team x Neutral Umpire present* is small and statistically insignificant.

In columns 5 and 6, we combine decisions by home and neutral umpires, and test whether the presence of a neutral colleague has a differential impact on home versus neutral umpires. The coefficient on the triple interaction *Foreign team x Neutral umpire present x Home umpire* indicates that being paired with a neutral colleague reduces a home umpire's decisions against foreign teams by 13-19% more than it impacts a neutral umpire.

These results provide compelling evidence that neutral peers have a debiasing effect on home

[19] We cannot identify the decision umpire for run out and stumping dismissals because the square leg umpire sometimes makes these decisions, and we cannot identify caught behind decisions because of missing fielder ID data for several matches in the decision-level data.

umpires. This debiasing effect is so large— almost 85% of the baseline level of in-group bias — that it largely eliminates bias against the foreign team.

A corollary of this finding is that introducing a second neutral umpire (i.e. requiring both on-field umpires to be neutral) may have limited scope to reduce bias further. We now test this implication.

4.3 The Limited Impact of a Second Neutral Umpire

To examine this hypothesis, we examine a second reform in 2002 that required Test matches to be refereed by two neutral umpires. We cannot use the same DiD design to estimate the impact of this reform, as a separate policy change affected ODI matches around the same time. Hence, we use an event study style design. Focusing on Test matches, we test whether foreign teams experienced fewer discretionary decisions relative to home teams after the 2002 reform. Specifically, we estimate the regression:

$$\begin{aligned} Y_{impgt} = \beta_1 Foreign_i + \beta_2 PostUmpiringReform_t \\ +\beta_3(Foreign_i \; x \; PostUmpiringReform_t) + \alpha_{pt} + \alpha_g + \alpha_j + X_m + \varepsilon_{impgt} \end{aligned} \tag{3}$$

where Y_{impgt} denotes discretionary decisions against team i in match m between teams p on ground g in year t. $Foreign_i$ is a dummy for decisions against the foreign team, $PostUmpiringReform_t$ indicates whether the match was held after the 1994 reform (which introduced 1 neutral umpire) or after the 2002 reform (which introduced the second neutral umpire). The coefficient of interest is β_3.

Panel A of Table 4 presents results. Column 1 replicates our baseline finding with a different specification (equation (3) rather than equation (1)), showing that the 1994 reform that introduced one neutral umpire reduced discretionary decisions against the foreign team by 8.6% (significant at 5% level). By contrast, column 2 shows that the 2002 reform which introduced a second neutral umpire had no impact, with a small coefficient (-2.6%) that is statistically insignificant (p=0.71). In column 3, we analyze the impact of both umpiring reforms in the same regression. The effect of the *first neutral umpire* reform remains similar (an 8% reduction in discretionary decisions against the foreign team) but now marginally loses statistical significance (p=0.14). However, the coefficient on the "second neutral umpire" reform becomes even smaller (-1.2%) and less statistically significant (p=0.86).

These results show that the first neutral umpire significantly reduced bias, in large part by debiasing the remaining home umpire. As a result, the second neutral umpire had little scope to reduce bias further. With strong debiasing effects, it may not be necessary to remove all conflicted evaluators to eliminate bias.

5 What Drives Debiasing? The Role of Social Pressure

We now explore the mechanisms behind debiasing effects. In a nutshell, our argument is that the physical presence of neutral peers exerts social pressure on (partisan or conflicted) evaluators to be impartial.

We first examine whether the debiasing effects of neutral colleagues are temporary or persistent. According to the contact hypothesis, inter-group contact leads people to moderate their beliefs and preferences, resulting in a sustained decrease in bias (Allport, Clark, and Pettigrew, 1954; Bazzi, Gaduh, Rothenberg, and Wong, 2019; Lowe, 2021). The neutral umpire reform caused umpires to have much more contact with people from other countries—fellow umpires, players and cricket officials as well as regular citizens. If these social interactions were responsible for debiasing, we should expect to see persistent changes in umpire behavior.

We test whether home umpires who co-referee a match with a neutral colleague continue to be less biased in subsequent matches (e.g. in the next 1-2 weeks). Restricting attention to decisions by home umpires, we estimate the regression:

$$\begin{aligned} Y_{ijmpgt} = \beta_1 Foreign_i + \beta_2 NeutralPresent_m + \beta_3 PriorMatchExposure_j + \\ \beta_4 (Foreign_i \ x \ NeutralPresent_m) + \beta_5 (Foreign_i \ x \ PriorNeutralPeer_j) + \alpha_{pt} + \alpha_g + \tau X_m + \varepsilon_{impgt} \end{aligned} \tag{4}$$

$NeutralPresent_m$ denotes whether a neutral umpire is present in the current match. $PriorMatchExposure_j$ denotes the home umpire's role in his most recent match.[20] The coefficients of interest are β_4, which captures the contemporaneous debiasing effect of a neutral colleague's presence, and β_5, which captures the effect of prior exposure to a neutral umpire.

Panel B of Table 5 presents results. Two patterns are noteworthy. First, in all 3 regressions, the contemporaneous effect of working with a neutral umpire is large and statistically significant: in the presence of a neutral colleague, home umpires give 18-22% fewer decisions against the foreign team. Moreover, the magnitude of the contemporaneous effect is generally unaffected by controlling for prior exposure.

Second, we consistently find weak effects of prior exposure. Working with a fellow home umpire in the previous match increases bias in the current match (column 1), but the coefficient is small (5%) and statistically insignificant (p=0.52). Similarly, we find no persistent impact of being paired with a neutral colleague (column 2) or working as a neutral umpire (column 3) in the last match. The prior exposure coefficients are jointly insignificant. Moreover, we can reject that they are equal to the contemporaneous effect at the 5% level in two of the three cases.[21]

[20] There are 3 possibilities for $PriorMatchExposure_j$: (i) home umpire with home peer, (ii) home umpire with neutral peer, and (iii) neutral umpire. We restrict attention to matches where the home umpire officiated another match within the past 14 days, though our results are robust to selecting other time horizons.

[21] The p-value for the χ^2 test for joint significance of the prior exposure coefficients is 0.21

Working with neutral colleagues thus appears to have only a *temporary* debiasing effect on home umpires. This casts doubt on a contact hypothesis-style explanation. Instead, it is more consistent with neutral umpire exerting social pressure on home umpires to be impartial. Alternatively, neutral peers might increase the salience of a home umpire's professional identity at the expense of his national identity, reducing in-group bias and triggering greater compliance with professional norms like impartiality.

5.1 Greater Debiasing when Home Umpires have Reputational Concerns

In this section, we present three pieces of evidence that the introduction of neutral umpires caused a larger decrease in bias when home umpires had strong social image or reputational concerns, suggesting an important role for social pressure.

We estimate whether the treatment effects of neutral umpires vary by umpire characteristics. Pooling home and foreign team innings, we estimate the following specification for sub-samples of our data with varying umpire characteristics:

$$\begin{aligned} Y_{impgt} = \beta_1 Foreign_i + \beta_2 Post_t + \beta_3 Test_m + \beta_4 (Foreign_i * Post_t) + \beta_5 (Foreign_i * Test_m) \\ + \beta_6 (Post_t * Test_m) + \beta_7 (Foreign_i * Test_m * Post_t) + \alpha_{pt} + \alpha_g + \tau X_m + \varepsilon_{impgt} \end{aligned} \tag{5}$$

The variable of interest is the triple difference *Foreign*Post*Test*, which captures the differential impact of the 1994 reform (that required Test matches to have one neutral umpire) on decisions against foreign vs home teams. We test whether this triple difference coefficient varies in separate sub-samples of data.

5.1.1 Neutral Umpires have a Larger Impact when Working with Biased Home Umpires

First, we show how the impact of neutral umpires varies with the home umpire's ex-ante level of in-group bias. We measure the bias of home umpires by estimating umpire fixed effects for all (617) umpires in our sample who refereed matches before the 1994 reform. Each umpire's FE identifies his average proclivity to give discretionary decisions favoring his own team.[22] We observe significant individual-level variation in bias levels among home umpires, with nationality and umpire cohort FEs together explaining only around 42% of variation.

Panel A of Table 6 presents results. A neutral umpire's presence reduces decisions against the foreign team by 41% (column 1, significant at 5% level) when the home umpire's pre-reform bias was above median. By contrast, neutral umpires have a weak effect (coefficient of +0.8%)

[22]The dependent variable in this regression is the difference between the number of discretionary decisions against the home vs the foreign team in a match. The umpire FE is thus a proxy for an umpire's average "contribution" to decisions favoring his own team across his career. We then construct percentiles for umpire bias, and explore heterogeneity by whether the home umpire was above or below the median level of bias.

when the home umpire has a track record of being less biased (column 2). Both coefficients are statistically different at the 1% level, indicating that neutral umpires have larger effects on biased home umpires.[23]

Highly biased home umpires may feel greater *social pressure* from neutral peers to moderate their behavior. Alternatively, biased home umpires may curb bias because of *career concerns*. For example, they may worry more about their bias being detected by cricketing authorities when working alongside a neutral colleague. We investigate career concerns more thoroughly in section 5.2, and provide evidence against it driving our results.

5.1.2 Senior Neutral Umpires have Stronger Impacts on Home Umpires

Second, we examine whether home umpires moderate their behavior more when paired with a senior neutral colleague. We estimate equation (5) splitting the sample by whether the neutral umpire had above or below median experience.[24] Panel A of Table 6 reports results. Foreign teams receive 21% fewer discretionary decisions when the neutral umpire is senior (column 3, significant at 5% level) but only 9% fewer decisions when the neutral umpire is less experienced (column 4, not significant). Once again, we can reject that these coefficients are equal.

Home umpires may be especially keen to appear impartial to senior colleagues because of social image concerns, i.e. they value senior colleagues' approval, even if this has no lasting material consequence; or career concerns, i.e. they worry that being seen as biased by senior colleagues could hurt their career prospects. We provide evidence against career concerns driving our results in section 5.2. A learning channel—in which home umpires adopt the habits and practices of senior colleagues—is inconsistent with our results because it would predict (i) neutral umpires have persistent effects and (ii) larger impacts for junior home umpires, neither of which we find.

5.1.3 Home Umpires are less Biased when Working with New Colleagues

First impressions matter, as the adage goes. We test whether umpires are less biased when paired with an umpire with whom they have not previously worked. We construct a *relationship length* variable equal to the number of matches a pair of umpires have refereed together. While the relationship length variable is (unsurprisingly) correlated with umpire experience ($\rho = 0.4$), there remains significant variation. Umpire experience, nationality, match location and year fixed effects together explain less than 40% of the underlying variation.

Columns 5 and 6 of Table 6, Panel A present results. Neutral umpires have a stronger effect when the on-field umpires have a new relationship, reducing decisions against the foreign team

[23] Our sample size for these regressions is lower because we only include matches that feature umpires who appear before and after the 1994 reform. We cannot estimate prior bias level for umpires who do not appear in matches before 1994.

[24] Umpire experience is measured by the number of matches they have officiated prior to the match in question.

by 21% (significant at the 5% level).[25] By contrast, neutral umpires have a more limited impact (coefficient of -6%, p=0.54) when the umpires have a more mature relationship. These coefficients are statistically different.

Our results show that neutral peers have stronger debiasing effects when the two umpires do not know each other well. If career concerns were driving this pattern, we should expect stronger effects for junior umpires (who are starting their careers) than experienced umpires (who have built a reputation). However, our results do not vary by umpire seniority, suggesting that they are better explained by social image concerns. Even experienced home umpires may be keen to make a good first impression with new colleagues.

5.2 External Monitoring Has a Limited Effect on Umpiring Bias

While the results above suggest that umpires do have reputational concerns when they are paired with colleagues in their line of sight, we now examine a separate reform that instead increased the external monitoring of umpires. In particular, the reform introduced match referees and TV umpires to all matches beginning in 1992. Figure 1b shows strong compliance with this reform. As described in section 2.4, match referees are tasked with ensuring fair play and high ethical standards. To this end, the match referee evaluates the performance of the on-field umpires. His match report awards each umpire a numerical performance score, and is submitted to the ICC for use in umpire promotion and contract renewal decisions. Thus, we consider the introduction of match referees a high-stakes performance monitoring reform that ought to trigger career concerns and affect umpire behavior.

We study how the match referee reform affected bias, by testing whether umpires gave fewer decisions against foreign teams after the reform. Specifically, we estimate the regression

$$Y_{impgt} = \beta_1 Foreign_i + \beta_2 PostMatchRef_t + \beta_3 (Foreign_i \; x \; PostMatchRef_t) + \alpha_{pt} + \alpha_g + \alpha_j + X_m + \varepsilon_{impgt} \tag{6}$$

where Y_{impgt} denotes discretionary decisions against team i in match m between teams p on ground g in year t. $Foreign_i$ is a dummy for decisions against the foreign team, $PostMatchRef_t$ indicates whether the match was held after the introduction of match referees and TV umpires. β_3 is the coefficient of interest.

Panel B of Table 4 presents results. In column 1, we estimate the regression equation restricting attention to ODI matches, which were unaffected by the neutral umpiring reform. We find that the introduction of match referees had no impact on bias (column 1), with a small point estimate (+2%) that is not close to statistical significance (p=0.73). Columns 2 and 3 report estimates from our main specification from Table 2 for the home team (col. 2) and the foreign team (col. 3) except they include a control for a match referee. The point estimates are virtually identical to our

[25] We split the sample into above- and below-median length umpiring relationships and examine the neutral umpire reform's impact on each subsample.

baseline estimates in table 2 (columns 1-2) and the statistical significance is unaffected, while the coefficient on the match referee dummy is small and statistically insignificant.

While the estimate of the effect of the monitoring reform in column 1 is somewhat imprecise, the finding is still substantially different—we can reject equality of coefficients with the neutral umpiring reform—from the neutral reform. Since match referees have explicit responsibility for monitoring and evaluating umpires' performance and ethical conduct, this finding lends support to the importance of physical presence in debiasing. Neutral umpires do not monitor or evaluate the home umpires they work with, but they stand in the direct line-of-sight of the home umpire. This may serve as a visual reminder to home umpires to be impartial (Ekström, 2012) and help to counteract crowd pressure. By contrast, while match referees have the power and responsibility to sanction biased umpires, and their views directly affect umpires' careers, they are not physically present on the field of play and typically observe the match from the spectator area. Their presence may thus be much less salient to the on-field umpires.

Our evidence points to the significant role that presence effects may play in mitigating in-group bias. Presence effects may also explain why peers influence bias in other contexts (e.g. juries, hiring committees), but it is usually not possible in these settings to distinguish presence effects from other mechanisms such as deliberation effects, strategic voting, career concerns, or inter-group contact channels. We make progress by exploiting features of, and natural experiments in, cricket to isolate presence effects.

5.3 Neutral Umpires are Less Affected by Social Pressure

In this section, we document an interesting asymmetry: unlike home umpires, neutral umpires are relatively unaffected by social pressure from crowds or peers. Earlier, we showed that home umpires are debiased by the presence of a neutral peer. By contrast, the behavior of neutral umpires is unaffected by whether they are paired with a home or neutral umpire (table 3, columns 3-4). They are unbiased in both cases, giving a similar number of decisions against home and foreign teams. Similarly, table 5 shows that, unlike home umpires, neutral umpires are not influenced by crowd pressure to give more decisions against foreign teams on weekend days (column 3).[26] These patterns hold with umpire fixed effects, implying that the same umpire is more resistant to social pressure when serving as a neutral umpire.

This might capture the effect of being given the (oxymoronic) title of *neutral* umpire. The label might increase the salience of an umpire's professional identity relative to other identities and encourage behaviors consistent with professional norms (such as impartiality and resisting social

[26] Appendix B5 provides further results on the effect of crowd pressure. Matches that begin on weekends see increased home umpire bias at the start of the match but not at the end, while matches that end of weekends see increased home umpire bias concentrated at the end of the match. Neutral umpires, however, show no such fluctuations in behavior and remain impartial throughout the match.

pressure).[27]

6 Robustness

6.1 Controlling for Umpire Characteristics

In this section, we deal with the concern that the umpiring reform may have reduced bias by changing the set of umpires who officiate international matches. To provide evidence against this, we re-estimate our baseline difference-in-difference regression with umpire fixed effects. Holding fixed umpire identity implies that any treatment effects of neutral umpires should be the result of changes in umpire behavior. Appendix table B3 shows that the baseline results are unchanged, with neutral umpires increasing discretionary decisions against home teams (column 1), decreasing discretionary decisions against foreign teams (column 2) and not affecting non-discretionary decisions. Indeed, the only noticeable difference with including umpire fixed effects is that we now see stronger evidence that the neutral umpire reform increased the odds that foreign teams win (column 6).

6.2 Player Responses to the Reform

As previously discussed, umpiring reforms can also affect player behavior. A home team batsman who believes neutral umpires make an LBW decision more likely may adjust his batting strategy to compensate. Likewise, home team bowlers may change their bowling strategy if they believe neutral umpires will be less favorable with discretionary decisions.

We estimate the effects of neutral umpires on proxies for player responses. Appendix table B4 presents results from equation (1), our baseline DiD specification. Panel B presents results for decisions where umpires have little discretion and thus changes likely reflect player responses. Columns 1-6 capture effects on different types of non-discretionary decisions; none of these 6 coefficients are statistically significant and 4 have a magnitude of less than 1%. Columns 7-8 show impacts on strike rate, a measure of how aggressively a batsman plays.[28] Neutral umpires had no impact on this proxy of playing style for both home and foreign team players.

Panel A shows effects on decisions that afford umpires more discretion. We see a clear pattern where the introduction of a neutral umpire results in home teams experiencing more discretionary decisions and foreign teams receiving fewer such decisions. The magnitudes are sizeable (larger than 6% for 5 of the 6 coefficients) and the coefficients are jointly significant. These results suggest that our baseline results are driven by changes in umpire behavior rather than player behavior.

[27] Prior work shows that individuals experience disutility if their behavior deviates from identity-prescribed norms (Cohn, Fehr, and Maréchal, 2014) which can be influenced by framing (Chang, Chen, and Krupka, 2019).

[28] The strike rate is the number of runs scored per 100 balls. It captures the rate of run-scoring and is usually used as a proxy for risk-taking and aggressive batsmanship.

6.3 External Validity

In order to provide support for the external validity of our findings, we test an important implication of our key mechanism (presence effects) in two disparate contexts. Presence effects imply that small changes to peer composition — e.g. the presence of one minority evaluator — can have large effects on bias. We re-analyze data from Neggers (2018), who studies in-group bias among election officials in India. Three election officers staff a polling booth, and make decisions about voter eligibility in each other's presence. The paper's main finding is that parties representing religious and ethnic minorities receive more votes at polling booths with minority election officials, because minority voters are less likely to be turned away.[29] Re-analyzing his data, we find that this effect is entirely driven by the presence of one minority election official on the three-person team (column 1 of Appendix table C1). As in our setting, a second minority official has no additional impact, even though this changes the balance of power at the polling booth. This pattern is consistent with majority-group officials curbing their in-group bias when in the presence of a minority-group colleague. However, unlike us, Neggers (2018) does not observe individual decisions and therefore cannot directly identify debiasing.

In a second example, we show evidence of peer influence on the US Court of Appeals. Reanalyzing data from Gelman, Liebman, West, and Kiss (2004), we show that conservative white judges are more likely to grant relief to capital punishment defendants when there is a single black or Hispanic judge on the panel.[30]

Presence effects can explain debiasing in both these settings, though one cannot rule out other plausible mechanisms, such as deliberation effects and dissent aversion.[31]

7 Conclusion & Discussion

In this paper, we studied how peers affect in-group bias among professional evaluators. Exploiting a series of umpiring reforms in cricket, we showed that partisan evaluators (i.e. home umpires) are less biased when working with a neutral colleague. This debiasing effect appears to be driven by the social pressure that home umpires feel to be impartial when in the physical presence of neutral peers. Such presence effects likely operate in many economic contexts, although it can be

[29] Caste and religion are the most salient group identities during Indian elections. The study was conducted in Bihar, India, and tested whether the RJD party (which represents Muslim and Yadav voters) received more votes when Muslim or Yadav election officials worked at the polling booth. Muslims and Yadavs together comprise approximately 33% of the population but are vastly under-represented in the bureaucracy, which is dominated by upper caste Hindus, who typically support another party, the BJP.

[30] Judges are effectively randomly assigned to cases, giving quasi-random exposure to minority group peers. Conservative white judges are more likely to vote to deny relief from capital punishment, but are less likely to do so when working with a minority group colleague. We classify judges as conservative if they have an above-median Judicial Common Space ideology score.

[31] In particular, judges may face incentives to coordinate their votes with their colleagues, since being the sole dissenter from the majority opinion often involves the cost of authoring a dissent note.

challenging to isolate them in settings (e.g. hiring committees, juries) where forces like deliberation and incentives to reach unanimity are often at work. Our findings highlight that physical presence may play an important role in debiasing and, more generally, in triggering social image concerns and mediating peer influence.

Our findings have several implications for the design of bias-mitigation policies. First, when peers exert strong debiasing effects, it may not be necessary to remove all conflicted evaluators to eliminate bias. This knowledge may be useful when there is a shortage of qualified evaluators and therefore potentially a tradeoff between bias and expertise.

Second, bias-mitigation policies that succeed in reducing discrimination along one axis may *inadvertently* increase disparities along other dimensions. In our context, professional umpires are disproportionately white. The neutral umpire mandates exacerbated this unequal representation, as home umpires were replaced by neutral umpires who were mostly white. In results we do not report in this paper, we find suggestive evidence that the neutral umpire reforms had a weaker effect on white home teams, because they were more likely to have a neutral umpire of their own race. Thus, while neutral umpires were successful in reducing in-group bias towards home teams, the policy may have unintentionally benefited white teams. This illustrates the complexity of designing bias-mitigation policies when identities are multi-dimensional or intersectional. Future research could explore this problem more systematically.

We also believe that it would be important to examine whether the effects of diversity initiatives are weakened if important decisions are increasingly taken at virtual meetings. Our findings suggest that our peers' physical presence plays an underappreciated role in mitigating bias; does a virtual presence have a similar effect? This ought to be a consideration for organizations when designing anti-bias policies in the context of increasingly common 'work from home' arrangements.

References

ALLPORT, G. W., K. CLARK, AND T. PETTIGREW (1954): "The nature of prejudice," . 5

ANWAR, S., P. BAYER, AND R. HJALMARSSON (2012): "The impact of jury race in criminal trials," *The Quarterly Journal of Economics*, 127(2), 1017–1055. 1, 1, 4

BAGUES, M., M. SYLOS-LABINI, AND N. ZINOVYEVA (2017): "Does the gender composition of scientific committees matter?," *American Economic Review*, 107(4), 1207–38. 1

BAGUES, M. F., AND B. ESTEVE-VOLART (2010): "Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment," *The Review of Economic Studies*, 77(4), 1301–1328. 1, 4

BARR, A., T. LANE, AND D. NOSENZO (2018): "On the social inappropriateness of discrimination," *Journal of Public Economics*, 164, 153–164. 1

BAZZI, S., A. GADUH, A. D. ROTHENBERG, AND M. WONG (2019): "Unity in diversity? how intergroup contact can foster nation building," *American Economic Review*, 109(11), 3978–4025. 1, 5

BERTRAND, M., AND E. DUFLO (2017): "Field Experiments on Discrimination," in *Handbook of Economic Field Experiments*, vol. 1, pp. 309–393. Elsevier. 1

BRODRIBB, G. (1999): *Next Man In: A Survey of Cricket Laws and Customs*. Souvenir. 7

BRYSON, A., P. DOLTON, J. J. READE, D. SCHREYER, AND C. SINGLETON (2021): "Causal effects of an absent crowd on performances and refereeing decisions during Covid-19," *Economics Letters*, 198, 109664. 3

BURSZTYN, L., G. EGOROV, AND S. FIORIN (2020): "From extreme to mainstream: The erosion of social norms," *American economic review*, 110(11), 3522–48. 1

BURSZTYN, L., AND R. JENSEN (2017): "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure," *Annual Review of Economics*, 9, 131–153. 1

CHANG, D., R. CHEN, AND E. KRUPKA (2019): "Rhetoric matters: A social norms explanation for the anomaly of framing," *Games and Economic Behavior*, 116, 158–178. 27

COHEN, A., Z. NEEMAN, AND F. AUFEROTH (2021): "Judging Under Public Pressure," Discussion paper, National Bureau of Economic Research. 3, 1

COHN, A., E. FEHR, AND M. A. MARÉCHAL (2014): "Business culture and dishonesty in the banking industry," *Nature*, 516(7529), 86–89. 27

DE PAOLA, M., AND V. SCOPPA (2015): "Gender discrimination and evaluatorsâ gender: evidence from Italian academia," *Economica*, 82(325), 162–188. 4

DEPETRIS-CHAUVIN, E., R. DURANTE, AND F. CAMPANTE (2020): "Building nations through shared experiences: Evidence from African football," *American Economic Review*, 110(5), 1572–1602. 2.2

DOHMEN, T., AND J. SAUERMANN (2016): "Referee bias," *Journal of Economic Surveys*, 30(4), 679–695. 1

EKSTRÖM, M. (2012): "Do watching eyes affect charitable giving? Evidence from a field experiment," *Experimental Economics*, 15(3), 530–546. 5.2

ENDRICH, M., AND T. GESCHE (2020): "Home-bias in referee decisions: Evidence from âGhost Matchesâ during the Covid19-Pandemic," *Economics Letters*, 197, 109621. 3

FEESS, E., H. MUELLER, ET AL. (2021): "Favoritism towards High-Status Clubs: Evidence from German Soccer," *The Journal of Law, Economics, and Organization*. 1

FLANAGAN, F. X. (2018): "Race, gender, and juries: Evidence from North Carolina," *The Journal of Law and Economics*, 61(2), 189–214. 4

GARICANO, L., I. PALACIOS-HUERTA, AND C. PRENDERGAST (2005): "Favoritism under social pressure," *Review of Economics and Statistics*, 87(2), 208–216. 3, 1

GELMAN, A., J. S. LIEBMAN, V. WEST, AND A. KISS (2004): "A broken system: the persistent patterns of reversals of death sentences in the United States," *Journal of Empirical Legal Studies*, 1(2), 209–261. 6.3, **??**

GROSSMAN, G., O. GAZAL-AYAL, S. D. PIMENTEL, AND J. M. WEINSTEIN (2016): "Descriptive representation and judicial outcomes in multiethnic societies," *American Journal of Political Science*, 60(1), 44–69. 4

GUERIN, B. (1986): "Mere presence effects in humans: A review," *Journal of experimental social psychology*, 22(1), 38–77. 1

GURYAN, J., K. KROFT, AND M. J. NOTOWIDIGDO (2009): "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments," *American Economic Journal: Applied Economics*, 1(4), 34–68. 1

HARRIS, D., B. HERRMANN, A. KONTOLEON, AND J. NEWTON (2015): "Is it a norm to favour your own group?," *Experimental Economics*, 18(3), 491–521. 1

HOEKSTRA, M., AND B. STREET (2021): "The Effect of Own-Gender Jurors on Conviction Rates," *The Journal of Law and Economics*, 64(3), 513–537. 4

LIST, J. A. (2004): "The nature and extent of discrimination in the marketplace: Evidence from the field," *The Quarterly Journal of Economics*, pp. 49–89. 1

LOWE, M. (2021): "Types of contact: A field experiment on collaborative and adversarial caste integration," *American Economic Review*, 111(6), 1807–44. 1, 5

MAS, A., AND E. MORETTI (2009): "Peers at work," *American Economic Review*, 99(1), 112–45. 1

NEGGERS, Y. (2018): "Enfranchising your own? experimental evidence on bureaucrat diversity and election bias in India," *American Economic Review*, 108(6), 1288–1321. 1, 6.3, **??**

PRICE, J., AND J. WOLFERS (2010a): "Racial discrimination among NBA referees," *The Quarterly Journal of Economics*, (November). 1

PRICE, J., AND J. WOLFERS (2010b): "Racial discrimination among NBA referees," *The Quarterly Journal of Economics*, 125(4), 1859–1887. 1, 3.1.2

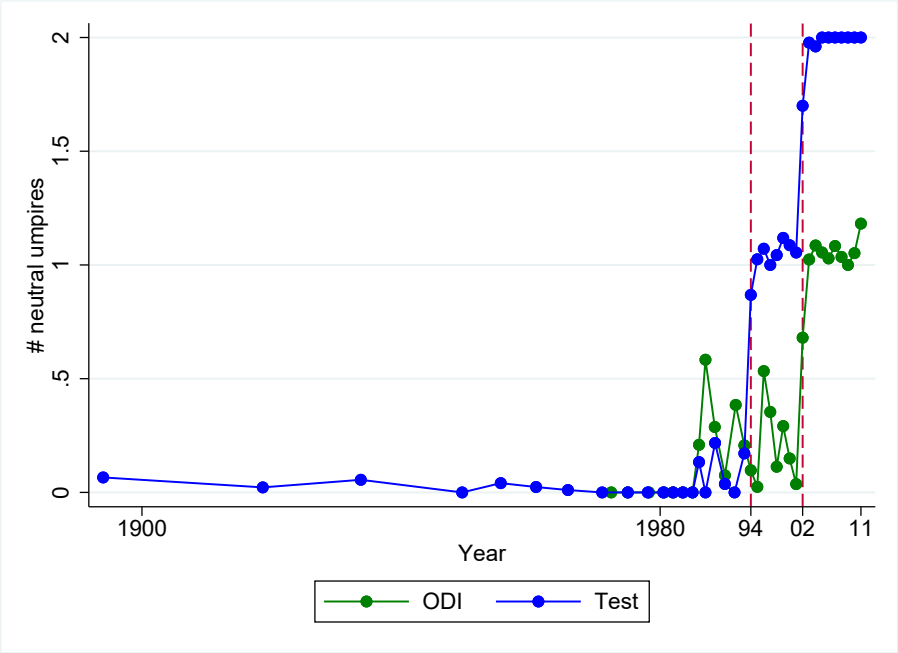
SACHETI, A., I. GREGORY-SMITH, AND D. PATON (2015): "Home bias in officiating: evidence from international cricket," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3), 741–755. 1, 3.1.2

SANDBERG, A. (2018): "Competing identities: a field study of in-group bias among professional evaluators," *The Economic Journal*, 128(613), 2131–2159. 1

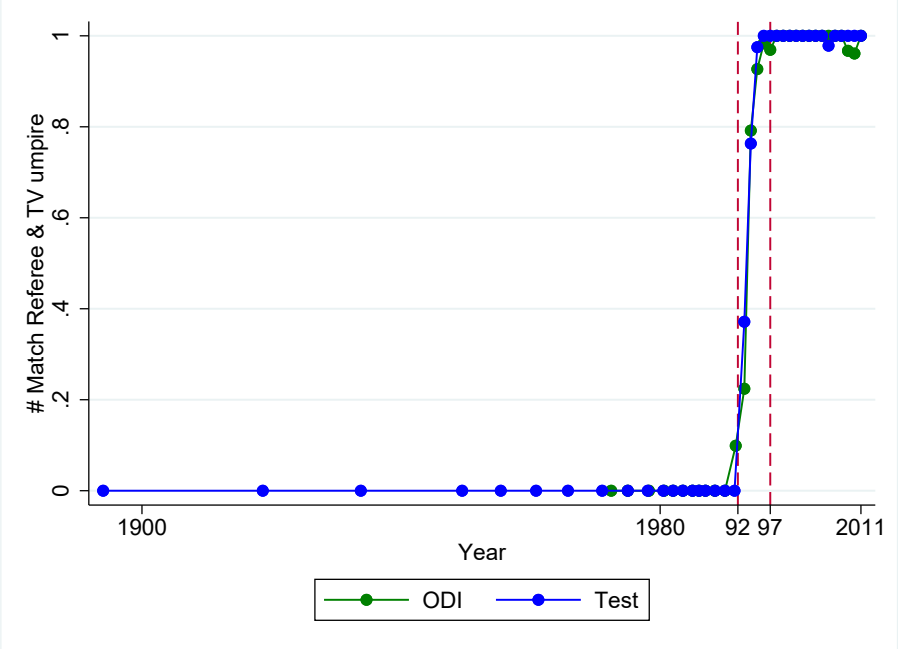
SHAYO, M. (2020): "Social identity and economic policy," *Annual Review of Economics*, 12, 355–389. 1

Figure 1: Reforms to Umpiring Panels

(a) Neutral Umpire reforms



(b) Introduction of Match Referees and TV umpires



The figure in Panel A plots the mean number of neutral umpires per match in a given year for Tests (blue line) and One Day Internationals (green line). In Panel A, the first vertical red dashed line signifies the year of the Test match neutral umpiring reform in 1994, while the second vertical red dashed line signifies the year in which two neutral umpires were mandated for Tests while one neutral umpire was mandated for ODIs. Panel B plots the fraction of matches in a given year that have a TV Umpire. The first vertical dashed line signifies the first year of adoption, while the second signifies universal adoption.

Table 1: Summary Statistics

	Mean	SD	#
Panel A: Dismissal characteristics			
LBWs	1.26	1.52	9586
Runouts	0.69	0.87	9586
Stumped	0.24	0.54	9586
Caught wicketkeeper	0.78	0.87	9482
Bowled	2.19	2.02	9586
Hit wicket	0.020	0.15	9586
Caught outfield	5.25	3.54	9482
Discretionary dismissals	3.00	2.09	9482
Non-Discretionary dismissals	7.48	4.54	9482
Panel B: Match characteristics			
Test match	0.40	0.49	5020
Home team wins	0.47	0.50	5020
Draw	0.14	0.35	5020
Foreign team wins	0.13	0.33	5020
Panel C: Umpire characteristics			
All matches			
Two home umpires	0.44	0.50	5020
One home, one neutral umpire	0.25	0.43	5020
One home, one away umpire	0.01	0.099	5020
Two neutral umpires	0.30	0.46	5020
Has TV umpire & match referee	0.57	0.50	5020
Experience (# matches umpired)	50.2	51.2	5020

Note: This table reports summary statistics. Panel A presents the number of dismissals of different categories for each team in a match. *Discretionary decisions* refer to LBW, runout, stumping and caught wicketkeeper decisions. Bowled, hit wicket and caught outfield are *non-discretionary decisions*. Panel B contains information on team characteristics and ultimate match outcomes. Panel C contains information on the composition of umpires in each match and describe umpire characteristics such as experience and race.

Table 2: The Impact of Introducing a Neutral Umpire (Team-Level Outcomes)

	IHS(discretionary decisions)		*IHS(non-disc. decisions)*		*Close match*	*Away win*
	(1)	(2)	(3)	(4)	(5)	(6)
Test * Post-1994	0.142**	-0.233***	0.004	-0.002	0.095*	0.066
	(0.071)	(0.071)	(0.066)	(0.074)	(0.052)	(0.048)
Sample (Inning)	Home	Foreign	Home	Foreign	All	All
N	2593	2593	2593	2593	2602	2602
Adjusted R^2	0.265	0.278	0.421	0.401	0.078	0.167
Depvar Mean	1.708	1.697	2.629	2.641	0.300	0.127

Note: This table presents results from our baseline DiD regression, specifically estimates of the key coefficient *Test * Post1994* , which is the interaction between *Test* (a dummy for Test matches) and *Post1994* (a dummy for matches held after the reform requiring Test matches to have 1 neutral umpire). In columns 1-2, the dependent variable is the inverse hyperbolic sine transformation (IHS) of the number of discretionary decisions against home (col 1) and foreign (col 2) teams. In column 3-4, the dependent variable is the IHS transformation of the number of non-discretionary decisions for home (col 3) and foreign (col 4) team innings. Discretionary decisions include LBWs, runouts, stumped, caught by the wicketkeeper, obstructing play, and handling the ball. Non-discretionary decisions include bowled, hit wicket, caught in the outfield, absent hurt, absent ill, retired hurt, retired out and retired not out. The dependent variable in Column 5 is a dummy indicating a close match, defined as a below median margin of victory, while in column 6 it is a dummy indicating a foreign team win. All regressions include Team Pair*Year FEs, ground FEs, and controls for whether a TV umpire and match referee were present. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: The Effect of a Neutral Umpire's Presence by Umpire Identity (Umpire-Level Outcomes)

	IHS (LBW dismissals)					
	(1)	(2)	(3)	(4)	(5)	(6)
Foreign Team	0.144***	0.145***	0.052	0.045	0.061**	-0.011
	(0.020)	(0.022)	(0.033)	(0.034)	(0.031)	(0.034)
Foreign team*Neutral ump present	-0.131**	-0.127**	0.031	0.028	0.020	0.095
	(0.055)	(0.064)	(0.058)	(0.062)	(0.057)	(0.062)
Foreign team*Home ump decision					0.083**	0.155***
					(0.037)	(0.040)
Foreign team*Neutral ump present*Home ump decision					-0.133*	-0.196**
					(0.079)	(0.084)
Sample (Decisions by umpire)	Home umpire		Neutral umpire		All	
Umpire FE	No	Yes	No	Yes	No	Yes
N	4930	4925	3267	3175	8277	8261
R2	0.218	0.245	0.188	0.197	0.200	0.282
Depvar Mean	0.791	0.791	0.709	0.709	0.757	0.757

Note: *Foreign team* is a dummy for innings of the away team. *Neutral umpire present* is a dummy indicating that the other umpire officiating the match is a neutral umpire. *Home ump* is a dummy indicating that the decision was made by a home umpire. The dependent variable in all columns is the inverse hyperbolic sine (IHS) transformation of the number of LBW decisions given by a particular umpire in a match. All regressions include fixed effects for the ground on which the match is played, Team Pair*Year fixed effects, and a control for whether a TV umpire and match referee were present. Additionally, umpire fixed effects are included in columns 2, 4 and 6. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The Impact of Additional Umpiring Reforms

	IHS (Discretionary decisions)		
	(1)	(2)	(3)
Panel A: Second neutral umpire			
Foreign team * Post one neutral ump rule	-0.086**		-0.080
	(0.043)		(0.055)
Foreign team * Post two neutral ump rule		-0.026	-0.012
		(0.071)	(0.068)
Sample (matches)	Tests	Tests after 1994	Tests
N	3956	1412	3956
Adj. R^2	0.113	0.082	0.113
Depvar Mean	1.666	1.582	1.656
Panel B: Monitoring by Match Referee			
Foreign team * Post match referee reform	0.020		
	(0.058)		
Post one neutral ump rule * Test match		0.142**	-0.233***
		(0.071)	(0.071)
Match referee & TV umpire		-0.007	-0.053
		(0.136)	(0.138)
Sample (Team innings)	All	Home	Foreign
N	3432	2593	2593
Adjusted R^2	0.056	0.265	0.278
Depvar Mean	1.353	1.353	1.353

Note: *Post One Neutral Ump* is a dummy for matches played after the 1994 rule mandating one neutral umpire in Test matches. *Foreign team* is a dummy for innings of the away team. *Post two neutral ump* is a dummy for matches after 31 Mar 2002, when the rule requiring both umpires in Test matches to be neutral came into effect. *Post match ref* is a dummy for matches occurring after the 1992 reform that introduced match referees and TV umpires. *Match ref & TV umpire* is a dummy for whether a particular match had a match referee and TV umpire. In all regressions, the dependent variable is the IHS transformation of the number of discretionary decisions against a team in a match. All regressions include Team Pair*Year and ground fixed effects, and a control for whether the home team won the toss. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The Role of Social Pressure in Umpire Debiasing

	IHS (LBW decisions)		
	(1)	(2)	(3)
Panel A: Crowd pressure			
Foreign team	0.085***	0.085***	0.017
	(0.022)	(0.022)	(0.029)
Foreign team * Weekend innings	0.140***	0.140***	0.062
	(0.053)	(0.053)	(0.058)
Foreign team*Weekend innings*Neutral present		-0.185***	
		(0.060)	
Sample (Umpire)	Home	Home	Neutral
Sample (Inning)	All	All	All
N	4208	2330	2696
Adj. R^2	0.123	0.186	0.145
Panel B: Transience of debiasing effect			
Foreign* Neutral ump present	-0.184*	-0.211**	-0.227***
	(0.106)	(0.092)	(0.083)
Foreign* Last match home ump, home peer	0.050		
	(0.077)		
Foreign* Last match home ump, neutral peer		-0.032	
		(0.098)	
Foreign* Last match neutral ump			0.011
			(0.052)
N	2809	2809	2809
p-value: $\beta_{contemporaneous} = \beta_{prior\ exposure}$	0.005	0.265	0.017
Adjusted R^2	0.233	0.231	0.231

Note: *Foreign team* is a dummy for away team innings. *Weekend innings* is a dummy for whether the innings took place on a weekend. *Neutral present* is a dummy for whether the other umpire in the match is a neutral umpire. *Last match [umpire type]* is a dummy indicating whether the home umpire served in his previous match as a (i) home umpire w/ another home umpire (col 1), (ii) home umpire w/ a neutral colleague (col 2) or (iii) neutral umpire (col 3). The dependent variable is the IHS transformation of the number of LBW decisions given by a particular umpire in a match. In Panel A, all regressions include match FEs, while in Panel B all regressions include Team Pair*Year and Ground FEs, and a control for whether a TV umpire and match referee were present. All regressions in Panel B only include decisions by home umpires. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Heterogeneous Impacts of the Neutral Umpiring Reform

	Outcome variable: **IHS (Discretionary dismissals)**					
	Initial bias of home umpire		Experience of neutral umpire		Relationship length	
	(1)	(2)	(3)	(4)	(5)	(6)
Post1994* Test* Foreign team	-0.418**	0.008	-0.218**	-0.099	-0.066	-0.211*
	(0.163)	(0.128)	(0.096)	(0.094)	(0.108)	(0.120)
Sample	$>$ *median*	$<$ *median*	$>$ *median*	$<$ *median*	$>$ *median*	$<$ *median*
N	2256	2127	4722	4930	3599	2836
Adjusted R^2	0.227	0.254	0.227	0.225	0.228	0.216
p-value: $\beta_{above\,median} = \beta_{below\,median}$	0.000		0.0019		0.0695	

Note: *Post-1994* is a dummy for matches played during and after 1994. *Test* is a dummy for Test Matches. *Foreign* is a dummy for the foreign team. The dependent variable in all columns is the IHS transformation of the number of discretionary decisions against a team in a match. Discretionary decisions include LBWs, runouts, stumped, obstructing play, and handling the ball. All regressions include fixed effects for the ground on which the match is played, Team Pair*Year fixed effects, and a control for whether a TV umpire and Match referee were present. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix

Appendix A1: Discretionary and Non-discretionary decisions

Non-discretionary decisions

- **Bowled.** A batsman is out *bowled* if he misses the ball and it hits the wickets.[32] Umpires have virtually no discretion with bowled dismissals. If the ball hits the wickets, the batsman is out bowled; if it does not hit the wickets, he is not out. Approximately 21% of batsman are out bowled.

- **Hit wicket.** A batsman is out *hit wicket* if he (inadvertently) hits his wicket with his body or bat. Like bowled dismissals, this offers the umpire no discretion. This is a relatively rare dismissal: only 0.2% of batsmen are out hit wicket.

- **Caught outfield.** A batsman is out *caught* when he hits the ball in the air and a fielder catches the ball before it hits the ground. The umpire's discretion here is limited to judging whether (a) the batsman did indeed hit the ball and (b) the ball was caught by the fielder before it hit the ground. This is usually unambiguous when the ball is caught by fielders in the outfield. *Caught outfield* is the most common dismissal in our sample, accounting for 52% of all wickets.

Discretionary decisions

- **Leg-Before-Wicket.** A batsman is out *LBW* if the ball hits the batsman's leg and the umpire judges that the ball would have hit the wicket had the leg not obstructed it. The umpire must judge whether (a) the ball did indeed hit the leg (ie. did not hit the bat) and (b) would have gone on to hit the wicket. Several other criteria also apply, giving the umpire significant discretion over LBW dismissals.[33] Appendix A2 provides an illustration of an LBW decision. LBWs account for 12% of wickets, and are perhaps the most contentious mode of dismissal in cricket. Indeed, according to Brodribb (1999), "no dismissal has produced so much argument as LBW; it has caused trouble from its earliest days".

- **Run out & Stumped.** A batsman is *run out* when she attempts a run and a fielder collects and throws the ball, hitting the wickets before the batsman has reached the crease at the opposite end of the pitch[34]. The umpire must judge whether the batsman reached the crease before the ball hits the wickets. Similarly, a batsman is out *stumped* when she leaves the crease after missing the ball and the stumps are broken by the wicketkeeper. Before the advent of

[32] The wickets are the three wooden poles that the batsman stands in front of and protects as he faces the bowler.

[33] For example, the umpire must judge whether the batsman is offering a shot or just leaving the ball. If he is offering a shot, the umpire must also judge whether the ball pitched in line with the wicket.

[34] The crease refers to the line in front of the wicket that denotes the batsman's safe zone.

video replay technology, umpires judged runouts and stumpings with the naked eye, and this involved considerable discretion. Together, these account for 9.5% of all dismissals.

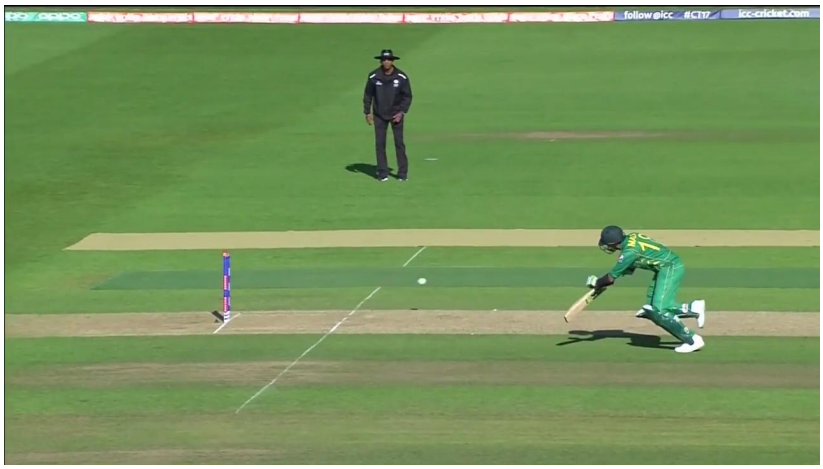
- **Caught wicketkeeper.** 5% of batsman are out caught by the wicketkeeper, who stands just behind the wickets. In these situations, the ball may lightly touch the bat and its trajectory may not noticeably deviate. It is often not clear whether the batsman has actually hit the ball, and thus there is significant scope for umpiring discretion.

Appendix A2: Common Dismissal Types

(a) The Batsman is 'Bowled'

(b) A 'Caught' Dismissal

(c) A Runout Decision

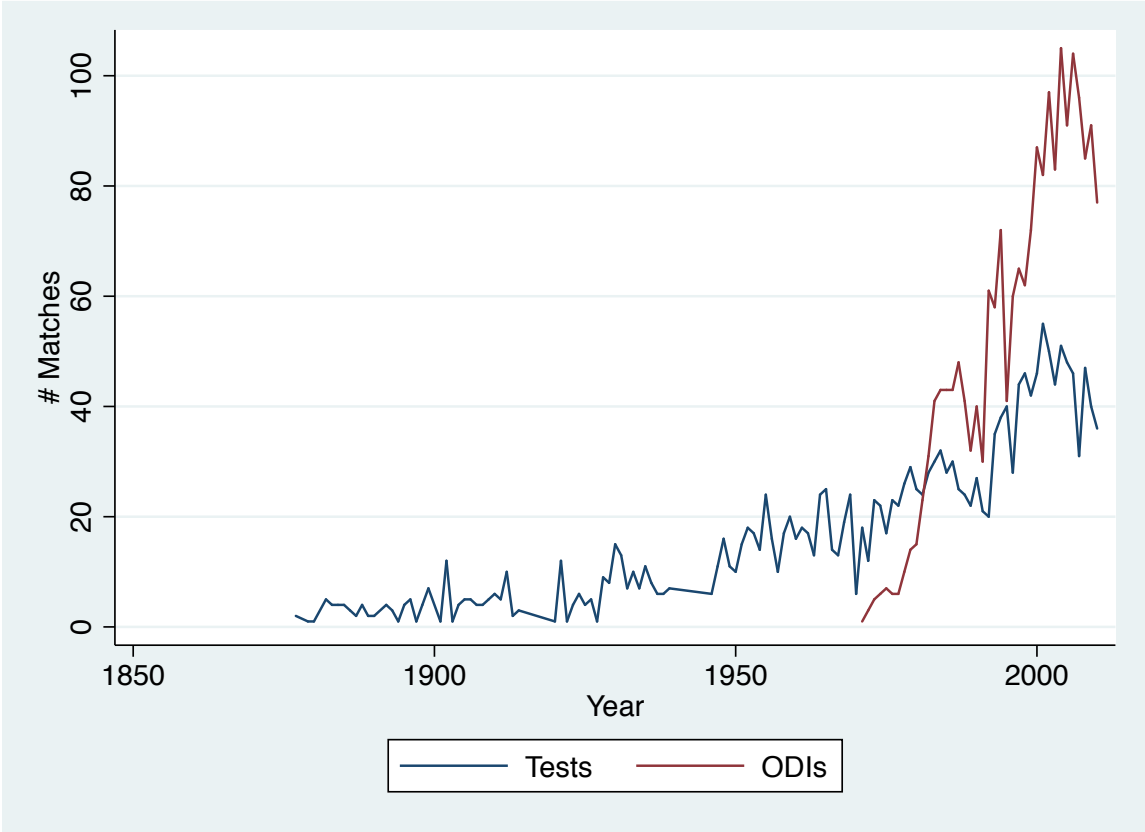


Sources: Channel 9 Australia and ESPNCricinfo.com

Appendix A2: The Leg Before Wicket (LBW) Dismissal

Note: The figure above displays a typical LBW scenario where the ball has hit the pads of the batsman and the umpire must make a judgment as to whether the ball will go on to hit the wicket. The pink line tracks the ball up to the point it hits the batsman's leg. The blue line shows the path of the ball were it not obstructed. The umpire must judge whether the unobstructed path of the ball would hit the wicket. If the primary umpire rules this is the case, the batsman is dismissed.
Source: Channel 9 Australia

Appendix A3: Number of Matches



Note: The chart above shows the number of matches for Tests and ODIs restricting attention to those not played on a neutral venue.

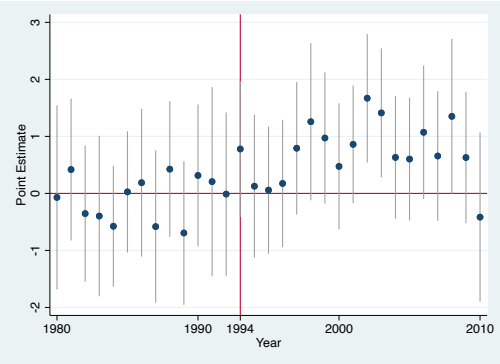
Appendix B1: Average LBWs by Umpire Type, Panel Composition, and Team Type

Umpiring panel	**Umpire**	**Team**		
		Foreign (1)	Home (2)	Difference (3)
Home-Home				
	Home	0.65	0.47	0.17***
		(0.83)	(0.72)	(0.03)
	# innings	1,072	1,068	
Home-Neutral				
	Home	0.46	0.39	0.07
		(0.66)	(0.65)	(0.06)
	# innings	241	223	
	Neutral	0.46	0.36	0.09
		(0.64)	(0.61)	(0.06)
	# innings	246	224	
Neutral-Neutral				
	Neutral	0.48	0.49	-0.01
		(0.72)	(0.76)	(0.08)
	# innings	182	182	

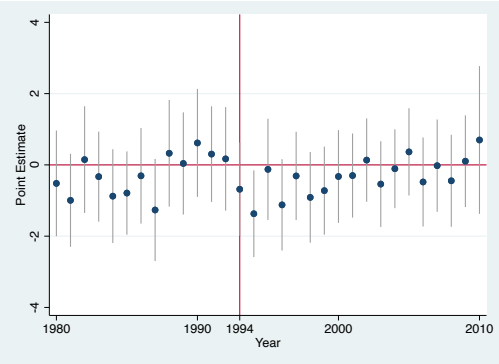
Note: This table uses umpire-inning level data from One Day Internationals (ODIs). Each value is the average number of LBW decisions by an umpire type (home or neutral) awarded against a specific team (home or foreign) in a specific umpiring panel (home-home, home-neutral, neutral-neutral). In columns 1 and 2, values in parentheses are standard deviations, in column 3, they are robust standard errors.

Appendix B2: Event Study for Discretionary Decisions

(a) Discretionary Decisions - Home



(b) Discretionary Decisions - Foreign



Note: Panel A plots the year by year coefficients in a difference-in-difference specification comparing total discretionary decisions in Tests vs. ODIs, for the home team, while Panel B plots the analogous coefficients for the foreign team. The year of reference in each case is 1993 and the vertical red line plots the year of the 1994 neutral umpiring reform in Tests. The null hypothesis for the joint F-test for the lags (1980-1993) cannot be rejected in each case. For our primary outcome, the inverse hyperbolic sine transformation of the number of discretionary dismissals, the joint F-test is 0.93 (p-value=0.52) for home team outcomes and 1.01 (p-value=0.44) for foreign team outcomes. When using the raw number of discretionary dismissals as the dependent variable, the joint F-test is 1.15 (p-value = 0.31) for home team outcomes and 1.27 (p-value = 0.23) for foreign team outcomes.

Appendix B3: Robustness of Reduced Form Results to Inclusion of Umpire Fixed Effects

	IHS(discretionary decisions)		*IHS(non-disc. decisions)*		*Close match*	*Away win*
	(1)	(2)	(3)	(4)	(5)	(6)
Test * Post-1994	0.234**	-0.202*	0.007	0.071	0.092	0.157**
	(0.097)	(0.105)	(0.094)	(0.106)	(0.078)	(0.071)
Sample (Inning)	Home	Foreign	Home	Foreign	All	All
Umpire FE	Yes	Yes	Yes	Yes	Yes	Yes
N	2254	2254	2254	2254	2263	2263
Adjusted R^2	0.294	0.262	0.392	0.384	0.015	0.151
Depvar Mean	1.708	1.697	2.629	2.641	0.300	0.127

Note: This table presents results from our baseline DiD regression, specifically estimates of the key coefficient *Test * Post1994* , which is the interaction between *Test* (a dummy for all Test matches) and *Post1994* (a dummy for all matches held after the reform requiring Test matches to have 1 neutral umpire). In columns 1-2, the dependent variable is the inverse hyperbolic sine transformation (IHS) of the number of discretionary decisions. In column 3-4, the dependent variable is the IHS transformation of the number of non-discretionary decisions for home (col 3) and foreign (col 4) team innings. Discretionary decisions include LBWs, runouts, stumped, caught by the wicketkeeper, obstructing play, and handling the ball. Non-discretionary decisions include bowled, hit wicket, caught in the outfield, absent hurt, absent ill, retired hurt, retired out and retired not out. The dependent variable in Column 5 is a dummy indicating a close match, defined as a below median margin of victory, while in column 6 it is a dummy indicating a foreign team win. All regressions include Team Pair*Year FEs, ground FEs, umpire FEs, and controls for whether a TV umpire and match referee were present. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B4: Umpire and Player Responses to Introduction of Neutral Umpires

Panel A: Discretionary Decisions	*IHS (LBW)*		*IHS (runout / stumped)*		*IHS (Caught wicketkeeper)*		*Away win*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test * Post-1994	0.232***	-0.061	0.081	-0.093	0.020	-0.145**	-0.033	0.172**
	(0.077)	(0.072)	(0.071)	(0.072)	(0.060)	(0.060)	(0.077)	(0.078)
Level of analysis	Innings	Innings	Innings	Innings	Innings	Innings	Match	Match
Sample	Home	Foreign	Home	Foreign	Home	Foreign	Home toss	Foreign toss
N	2602	2602	2602	2602	2593	2593	1197	1143
Adjusted R^2	0.311	0.358	0.074	0.106	0.275	0.307	0.118	0.160
Depvar Mean	0.953	0.953	0.711	0.711	0.620	0.620	0.079	0.310
Panel B: Non-Disc. Decisions	*IHS (bowled)*		*IHS (hit wicket)*		*IHS (caught outfield)*		*IHS(Strike rate)*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test * Post-1994	-0.018	-0.087	0.001	-0.006	0.009	0.008	0.046	0.015
	(0.081)	(0.078)	(0.014)	(0.010)	(0.072)	(0.082)	(0.065)	(0.075)
Sample (Inning)	Home	Foreign	Home	Foreign	Home	Foreign	Home	Foreign
N	2602	2602	2602	2602	2593	2593	2593	2593
Adjusted R^2	0.324	0.306	0.126	0.096	0.364	0.350	0.831	0.838
Depvar Mean	1.594	1.594	0.035	0.035	2.366	2.366	3.130	3.130

Note: This table presents results from our baseline DiD regression, specifically estimates of the key coefficient *Test * Post1994* , the interaction between *Test* (a dummy for all Test matches) and *Post1994* (a dummy for all matches held after 1994, the year Test matches were required to have 1 neutral umpire). In panel A, the dependent variable is the IHS transformation of the number of LBW decisions (columns 1-2), runout and stumping decisions (columns 3-4) and caught by wicketkeeper decisions (columns 5-6). In column 7-8, the dependent variable is a dummy indicating a foreign team win. In panel B, the dependent variable in columns 1-6 is the IHS transformation of bowled, hit wicket and caught outfield decisions respectively, while in column 7-8 it is the batsman's strike rate, which measures how many runs are scored by 100 balls, a proxy for the aggressiveness of a batsman's playing style. All regressions include Team Pair*Year FEs, ground FEs, and controls for whether a TV umpire and match referee were present. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B5: Additional results on umpires' response to crowd pressure and officiating rival teams

	IHS (LBW dismissals)					
	Crowd pressure				Rival teams	
	(1)	(2)	(3)	(4)	(5)	(6)
Foreign Team*Match starts on weekend	0.146**	0.068	0.034	-0.165		
	(0.061)	(0.083)	(0.088)	(0.127)		
Foreign team*Match ends on weekend	0.068	0.198	-0.019	0.010		
	(0.090)	(0.125)	(0.075)	(0.102)		
Foreign team					0.021	0.024
					(0.045)	(0.055)
Sample (Decisions by umpire)	Home umpire		Neutral umpire		Neutral umpire	
Period of match	Start	End	Start	End	Whole match	
Sample (neutral umpire rival)		No restrictions			Home team	Foreign team
N	2330	1508	1500	1110	8277	8261
R2	0.186	0.036	0.158	0.051	0.200	0.282

Note: *Foreign team* is a dummy for innings of the away team. *Match starts on weekend* is a dummy for whether the match started on a weekend, while *Match ends on weekend* is a dummy for whether the match ends on a weekend. In columns 1 and 3, we restrict attention to the first two innings of a Test match, while in columns 2 and 4, we restrict attention to the final two innings of a Test match. Column 5 restricts attention to matches where the home team is a rival of the neutral umpire's nation, while column 6 restricts attention to matches where the foreign team is a rival of the neutral umpire's nation. The dependent variable in all columns is the IHS transformation of the number of LBW decisions given by a particular umpire in a match. In columns 1-4, regressions include match fixed effects, while in columns 5-6, regressions include Team Pair*Year and ground fixed effects. All regressions include a control for whether a TV umpire and match referee were present. Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B6: The Impact of Neutral Umpire Mandates on Umpire and Decision Quality

	% Elite umpires ineligible (1)	*Umpire quality* (2)	*Frac. correct decisions* (3)
One neutral umpire	0.090***		
	(0.005)		
Two neutral umpires	0.166***		
	(0.005)		
Share of elite umpires conflicted		0.070	-0.006
		(0.085)	(0.083)
Sample (Inning)	Home	Foreign	Home
N	8651	1920	1921
Adjusted R^2	0.649	0.040	0.030
Depvar Mean	0.109	0.503	0.736

Note: In columns 1, the dependent variable is *Frac Umpires Conflicted*, which is the share of elite umpires who are ineligible to referee a match due to nationality conflicts arising from the neutral umpire requirements. In column 3, the dependent variable is *Frac. Correct decisions*, which is the share of decisions reviewed in a match that the on-field umpire got correct. We are able to identify correct decisions by observing whether the umpire's original decision was upheld or reversed by the TV umpire during the review. In column 2, the dependent variable is *Umpire Quality*, which is the percentile rank (scaled between 0 and 1) of the umpire's estimated fixed effect. We estimate umpire FEs from a regression of *Frac. Correct decisions* on umpire dummies. All regressions include Home Team and Year FEs and controls for the match format (eg. Test or ODI match). Standard errors, clustered at the match level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C1: Testing Implications of Presence Effects in Other Settings

	RJD-BJP vote margin	Capital relief granted
	(1)	(2)
Exactly 1 Muslim/Yadav officer	0.034***	
	(0.012)	
Exactly 2 Muslim/Yadav officer	-0.030	
	(0.030)	
Exactly 3 Muslim/Yadav officer	-0.137	
	(0.248)	
Muslim/Yadav electors pct	1.481***	
	(0.030)	
Log total registered voters	-0.060***	
	(0.021)	
Conservative		-0.096***
		(0.026)
Conservative × minority judge on panel		0.104*
		(0.060)
N	5089	1450
Adjusted R^2	0.41	0.824
Depvar Mean	-0.18	0.364

Note: Column 1 presents regressions based on data from Neggers (2018). The dependent variable is the victory margin of the RJD party (largely supported by the minority Muslim and Yadav communities) relative to the BJP party (largely supported by upper caste Hindu communities). The key regressors are the number of Muslim/Yadav polling booth officers at each polling station. Standard errors are clustered at the polling station level. Column 2 uses data from Gelman, Liebman, West, and Kiss (2004) on capital relief cases heard by the US Court of Appeals. The dependent variable is whether an individual judge votes to grant relief to the defendant. The *Conservative* dummy captures whether a judge's ideology score is above the median level of conservativeness. The number of minority judges captures whether there was a Black and Hispanic judges on the three-judge panel. All regressions include individual judge, case and circuit fixed effects. Standard errors, clustered at the case level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$