Affiliation Bias in Arbitration: An Experimental Approach

Sergio Puig and Anton Strezhnev

ABSTRACT
A characteristic feature of arbitration, a growing form of legal adjudication, is that each disputing party appoints an arbitrator. Commentators, however, suggest that party-appointed arbitrators tend to be biased in favor of their appointers. Evaluating this claim from data on historical disputes is problematic because of nonrandom selection of arbitrators. Here we use a novel experimental approach to estimate the causal effect of the appointing party. Using survey experiments with arbitration experts around the world, we show that professional arbitrators suffer from affiliation effects—a cognitive predisposition to favor the appointing party. At a methodological level, we offer a solution to the problem of measuring this effect when credible observational designs are lacking.

1. INTRODUCTION
Arbitration has crept into nearly every corner of American life. Some scholars refer to this expansion as a “privatization of the justice system” (Silver-Greenberg and Corkery 2015, p. 1). This trend goes beyond domestic legal disputes, as arbitration has become a central element of global economic interdependence. In particular, the growth of arbitration has put arbitration panels in a position to rule on transnational investments and key political questions like the ability of governments to tax their citizens or regulate health—long areas of sovereign prerogative.

What makes this expansion potentially troubling is that systematic

Sergio Puig is an Associate Professor of Law in the James E. Rogers College of Law, University of Arizona. Anton Strezhnev is a Ph.D. candidate in the Department of Government, Harvard University. The authors wish to thank Christopher T. Robertson, Dave Marcus, Tom Ginsburg, Alan Sykes, Jonathan Greenberg, Tomer Broude, Alfred Yen, Christopher R. Drahozal, Susan Franck, Joshua Kertzer, and Beth A. Simmons and the peer reviewers and editors of the Journal of Legal Studies for excellent substantive and methodological feedback. Data are available at http://dx.doi.org/10.7910/DVN/9FWNBM.
features of arbitration make it difficult for arbitrators to be entirely un-biased in their decisions. Unfettered by precedent, deprived of strict uniform rules against conflicts of interest, and insulated from any judicial system, powerful corporations may use arbitration to steer cases to sympathetic arbitrators incentivized by the prospects of sizeable earnings (Branson 2010; Carter 2000).

This debate is not new in the legal academy (see, for example, Resnik 2004). What is new, however, is that more recent critiques of arbitration come from insiders who otherwise defend this dispute-settlement system as an efficient form of legal adjudication (see, for example, Paulsson 2010; Van den Berg 2010). These criticisms have focused on a particularly well-established feature of arbitration—party appointments, or the right of disputants to appoint an arbitrator to a tribunal. While such unilateral party appointments are well established, some now criticize the practice as a source of bias that negatively affects the impartiality and the legitimacy of arbitration proceedings.

Proposals for reform range from eliminating party appointments altogether to introducing changes in the selection of arbitrators, but most seem farfetched or unlikely as a practical matter (see, for example, Giorggetti 2013). In short, arbitration is designed to give parties greater control over the dispute-resolution process relative to a more formal judicial setting, and parties would be hesitant to give up one of the primary tools for exercising that control—the ability to (partially) choose who hears the dispute. However, a recent proposal that may be more plausible would permit parties to appoint arbitrators but would prevent nominees from knowing which party appointed them (American Bar Association 2013).

The use of blind appointments attempts to relieve the arbitrators of the possible affiliation or allegiance effects resulting from the nomination process but maintain unilaterally appointed arbitrators, which is considered a fundamental right in arbitration (Brower and Rosenberg 2013). The proposal is inspired by ideas ingrained in most legal systems, grounded in theories of justice (see, for example, Rawls 1985), and applied to the highest standards in scientific research—with the general principle being that impartiality can be enhanced by preventing a decision maker from accessing knowledge that might unduly affect the decision.

In this article, we use the blind-appointments proposal as an entry point to the debate over implicit biases of legal actors. First, we argue that an important issue in the debate over bias is to distinguish between
two mechanisms through which the nomination process affects arbitrators’ behavior: selection effects and affiliation effects. Prior observational research has shown that arbitrators tend to favor their appointing parties. However, because parties choose arbitrators in part on the basis of how they expect the arbitrator to vote, appointment-driven bias cannot be inferred from these correlations alone. To infer evidence of an affiliation effect (the implicit bias of the arbitrator to favor the appointing party), we need to eliminate the selection effect (the strategic decision of the litigant to choose arbitrators who are likely to be favorable). Therefore, we develop a set of randomized survey experiments designed to isolate affiliation effects in which we directly controlled arbitrators’ assigned appointers.

Second, we discuss the results of these experiments. In short, we presented surveyed arbitrators with a hypothetical choice task related to an investment arbitration case. Participants were randomly told that they were appointed by one of the parties, jointly or by agreement of the parties, or simply that they were appointed (without any information about the identity of the appointer). Using a sample of 257 arbitrators around the world, we found that arbitrators nominated by one of the two parties to the litigation tended to make decisions more favorable to that party compared with arbitrators appointed by the opposite party. We replicated this result in a follow-up experiment with a sample of 248 arbitrators. In addition, we found that arbitrators in the blind-appointment treatment exhibited similar responses to joint-party appointees and tended to take positions in the middle of the two parties.

Third, on the basis of these results, we argue that blinding could be a useful approach to reducing bias. The evidence we find strongly suggests that affiliation effects exist, and while there are certainly difficulties in implementing such a proposal, we suggest that blinding could help ameliorate affiliation biases while retaining the potential benefits of the parties’ involvement in the appointment process.

The article proceeds as follows: Section 2 provides some background on arbitration and the debate over party appointments. Section 3 describes the design of the initial experiment and its replication and reports the findings of both. Section 4 discusses some of the limitations of our study and practical issues with blinding arbitrators. Section 5 briefly concludes.
2. BACKGROUND

2.1. Unilateral Party Appointments: A Burgeoning Debate

As noted by Giorgetti (2013), a debate has emerged over unilateral party appointments. Critics of this system argue that the power and legitimacy of arbitration stem from an appearance of expertise, neutrality, and impartiality. Hence, the predisposition toward one party that results from the system of appointments poses significant challenges to the main legitimating aspects of arbitration. It may also result in unnecessary antagonism, complicated compromises, and inconsistent decisions in the awards of tribunals, all of which undermine the goal of an independent, rule-based adjudicatory system.

Defenders of the use of party-appointed arbitrators, on the other hand, argue that the traditional party-appointment system is the keystone of arbitration because it gives the parties some ownership over the process. In addition, when tribunals render unanimous decisions, the presence of party appointees may enhance the credibility of the award in the eyes of both parties since each litigant knows that a trusted appointee was willing to endorse the outcome. Being able to appoint an arbitrator, some argue, is one of the “most attractive aspects of arbitration as an alternative to domestic litigation” (Brower and Rosenberg 2013, p. 9), and its elimination would constitute a radical transformation, and potential devaluation, of arbitration.

The debate over the convenience of party appointments has been conducted almost exclusively with anecdotal evidence. However, a recent study using observational data demonstrates some of the merits of each side of this debate. By comparing cases with and without party appointments, Puig (2016) shows that tribunals without party-appointed arbitrators tend to handle cases faster and settle more often. Their decisions tend to be unanimous and are less likely to be questioned in a subsequent proceeding. Nonetheless, the rate of challenges to and resignations of arbitrators in tribunals without party-appointed arbitrators appears to be higher, and the higher rate of resignations in tribunals without party-appointed arbitrators tends to reduce the expediency of the process.

The debate has also shown a lack of consensus among arbitration professionals for moving away from the current system for the formation of tribunals. An international arbitration survey (Friedland and Brekoulakis 2012, p. 5) shows a general disapproval of “proposals calling for an end to unilateral party appointments.” In practice, this consen-
sus limits which reforms are feasible. Blind appointments, however, may resolve some of the concerns that emanate from the party-appointment system, without eradicating the practice of party appointments. Under a blind-appointment system, parties to an arbitration can continue to appoint arbitrators, but a mechanism is introduced to ensure that nominees do not know who appointed them. This proposal maintains a fundamental feature of party control but may help to mitigate the implicit bias and adversarial influence of explicitly known party appointments.

2.2. Party Affiliation: Arbitrators and Implicit Biases

Robertson (2010a, 2010b) demonstrates a number of different ways through which litigants can introduce biases in a legal process. When rules permit litigants to nominate a legal actor, they can ensure that such a person is not too independent minded by selecting someone who has shown reliability and the appropriate decision-making philosophy toward the relevant set of issues. In the arbitration context, the nomination and appointment of arbitrators by the parties is an important and calculated decision by the litigants.

As explained by Hunter (1986, p. 189), “when I am representing a client in an arbitration, what I am really looking for in a party-nominated arbitrator is someone with the maximum predisposition towards my client, but with the minimum appearance of bias.” Hence, the influence in the process that results from such practice could be attributed mainly to a selection effect.

Yet, even in the absence of a selection effect, appointees may also find it difficult to maintain impartiality because of implicit preferences for their appointing party or affiliation effects. In general, modern psychology research has extensively documented the existence of implicit biases in a variety of settings, from attitudes toward historically disadvantaged groups to the effect of primes and context cues on expressed beliefs and actions (see, for example, Greenwald and Banaji 1995). These implicit effects on attitudes often operate outside conscious cognitive modes—those expressing implicit biases may be entirely unaware of and unable to account for their influence. In particular, arbitrators may, despite their best intentions to remain unbiased, be nonetheless primed to favor their nominating party simply by knowing that they were selected by that party.

As suggested by Robertson and Kesselheim (2016), blinding, or removing key information that may affect a decision, is increasingly seen as an effective debiasing strategy. In part, blinding is a preferred intervention because the blind spot tends to persist even after individuals think
through and consider their biases. In fact, experimental research has shown a tendency for people to acknowledge bias more readily in others than in themselves, often increasing polarization in adversarial contexts. Hence, biases typically operate unconsciously, thereby leaving their influence hidden from direct introspection (Wilson and Brekke 1994).

3. EVIDENCE

3.1. Context

With some exceptions, arbitration tribunals are typically composed of three members—two party-appointed arbitrators and a third arbitrator, usually the chair, appointed by one of the following three methods: by agreement of the parties in the proceedings, by agreement of the two party-appointed arbitrators, or by an independent designating authority (commonly, the institution administering the proceedings). All arbitrators are supposed to be independent and impartial, including the two party-appointed arbitrators, and will often sign a declaration affirming their independence and impartiality.

Existing observational research on arbitral proceedings, such as analysis of dissent rates, is insufficient to differentiate between selection and affiliation effects. Moreover, little attention has been given to implicit biases in arbitral decision making (Drahozal 2004; Keer and Naimark 2001). To our knowledge, no prior experiment of this nature exists using arbitrators as experimental subjects.

3.2. Experiment

3.2.1. Design. Our pair of survey experiments are both based around a brief vignette describing a hypothetical arbitration between an investor and a state. Elements of the vignette were randomized for each participant to study how arbitrators’ responses might vary across different scenarios. Our primary manipulation of interest is the arbitrator’s appointing party. Participants were told that they were appointed by the respondent, by the claimant, or by agreement of the litigating parties or simply that they were appointed to the tribunal, with no mention of any appointing method. This last condition is what we refer to as the blind-appointment case.1 Appendix A provides a sample of the vignette that we presented to

1. Note that the surveyed arbitrators were not explicitly told that the identity of their appointing party was hidden from them. The text of the treatment simply reads, “You
arbitrators in the first experiment, with the key manipulation of interest shown in bold font.

In the first experiment, the vignette included a few additional manipulations to examine and control for other potential influences on arbitrators' behavior. We randomly varied the implied resource endowments of the claimant and the respondent. The claimant in the arbitration was a firm headquartered in either a high- or middle-income economy, while the respondent state was either a middle- or low-income economy. We also varied the type of ruling across four possible conditions reflecting different ways in which a case litigated under the applicable procedural rules in the vignette could be decided: the respondent expropriated the claimant's property (claimant wins), the respondent did not expropriate the claimant’s property (respondent wins), the dispute is outside of the tribunal’s jurisdiction (respondent wins on technical grounds), or the claimant’s claims are manifestly without legal merit (the case is dismissed summarily; that is, the respondent decisively wins). Each of the conditions had equal probability of being assigned to any individual, and the treatments were each randomized independent of one another. Table 1 summarizes the marginal distributions of the number of observations in each treatment condition in the sample.2

After being presented with the vignette, surveyed arbitrators were then asked how they thought the parties' expenses in the dispute, including the cost of legal representation, should be apportioned in such a case. Arbitrators could choose to have one party reimburse the other for all or some of its expenses or have each party bear its own expenses. Participants were then asked in an open-ended question to briefly discuss the reasoning behind their decision.

In the first vignette, we chose to ask the surveyed arbitrators to rule on

---

2. We present the marginal distributions of counts (rather than the full factorial joint distribution) as we are primarily interested in the marginal effects of each variable averaged over the distribution of the other treatments. While we do not have enough power to credibly estimate the effect of a particular unique combination of all four treatments, this is not the quantity in which we are interested. We do have sufficient observations to estimate marginal effects for a single treatment. In this sense, our experiment is very similar to the conjoint multiattribute choice experiment that is increasingly common in social science surveys. See Hainmueller, Hopkins, and Yamamoto (2014).

This content downloaded from 150.135.118.031 on January 29, 2018 09:51:47 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).
the allocation of costs rather than on the merits of a case for two reasons. First, summarizing the arguments of the parties such that participants would have enough information to render an educated decision on the merits of a full case would require an impossibly long vignette. We were conscious of the fact that participants (primarily busy business lawyers) would likely not have the time or interest to spend hours on our survey. Prompting arbitrators to render a clear decision on costs given a known result allowed us to use a vignette that participants could easily read and analyze in a practical amount of time.

Second, we needed an outcome that would generate variation in responses. Had we chosen an outcome on which there is obvious and clear legal guidance, we would expect arbitrators to all reach more or less the same result. In the context of our vignette, costs are an ideal outcome: the arbitration rules (ICSID 2006) grant complete discretion, and there is little precedential guidance for how arbitrators should use such discretion and allocate costs. Hence, some tribunals follow the principle that costs follow the event, and the losing party should compensate the winner for legal fees. Others choose to have each party bear its own costs. While the former appears to be the most common approach (see, for example, Table 1.

### Summary of Conditions in Experiment 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appointer:</strong></td>
<td></td>
</tr>
<tr>
<td>Appointed by the respondent</td>
<td>77</td>
</tr>
<tr>
<td>Appointed by the claimant</td>
<td>50</td>
</tr>
<tr>
<td>Joint appointment</td>
<td>67</td>
</tr>
<tr>
<td>Blind appointment</td>
<td>63</td>
</tr>
<tr>
<td><strong>Claimant’s wealth:</strong></td>
<td></td>
</tr>
<tr>
<td>Firm is from a high-income economy</td>
<td>135</td>
</tr>
<tr>
<td>Firm is from a middle-income economy</td>
<td>122</td>
</tr>
<tr>
<td><strong>Respondent’s wealth:</strong></td>
<td></td>
</tr>
<tr>
<td>Respondent is a middle-income economy</td>
<td>121</td>
</tr>
<tr>
<td>Respondent is a low-income economy</td>
<td>136</td>
</tr>
<tr>
<td><strong>Dispute outcome:</strong></td>
<td></td>
</tr>
<tr>
<td>Claimant wins on merits</td>
<td>60</td>
</tr>
<tr>
<td>Respondent wins on merits</td>
<td>66</td>
</tr>
<tr>
<td>Respondent wins on jurisdiction</td>
<td>73</td>
</tr>
<tr>
<td>Claimant is “manifestly without merit”</td>
<td>58</td>
</tr>
</tbody>
</table>

**Note.** Chi-squared tests for the marginal counts across all four conditions fail to reject the null hypothesis that the counts are generated by discrete uniform distributions ($p > .10$). $N = 257$ arbitrators.
International Chamber of Commerce 2015), it is rare that costs are fully borne by one party. There is a meaningful amount of open-endedness in the decision such that it could plausibly be swayed by extralegal considerations.

We conducted the first survey in the fall of 2015 and recruited participants by collecting publicly available e-mail addresses of arbitrators and lawyers who specialize in arbitration (the arbitrators). These addresses were largely obtained from lists published on the websites of arbitration institutions and in the directories of other professional organizations. We chose to recruit from this set of potential arbitrators, rather than from a typical sample of individuals from the general population, to make our results as generalizable as possible to the population of potential arbitrators around the world. Moreover, understanding and competently completing the vignette requires some familiarity with international arbitration, which makes typical samples of convenience useless. While not all participants indicated that they had served as arbitrators (in a survey administered after the vignette), all indicated some expertise in arbitration, in either investor-state or commercial arbitration. To increase our response rate, we sent several reminders to participants but promised only advance circulation of any articles summarizing the research as the incentive. We obtained approval from the institutional review boards of both the University of Arizona and Harvard University.

3.2.2. Analysis and Results. A total of 538 individuals responded to the e-mailed survey, but not all arbitrators completed the survey or answered all of the relevant questions. Therefore, we received only 257 complete responses to the investor-state vignette. While this may appear to be a very low response rate, this does not invalidate our experimental design. Random assignment of treatments allows us to obtain an unbiased estimate of the treatment effect for arbitrators regardless of how individuals were selected into the sample. Moreover, if the factors driving non-response are not associated with individuals’ effect sizes, we can treat our effect estimates in the sample as representative of the average effect in the population. Nevertheless, we thoroughly discuss the generalizability of the results in Section 3.3, which examines our sample composition.

In allocating costs, arbitrators are first and foremost driven by the out-

3. In this experiment, we sent out 28,832 recruitment e-mails. However, some of them were returned with a delivery failure message, as the addresses were no longer active. Because arbitrators could choose to stop the survey at any time, we could not force them to respond to all questions in the survey.
come of the case. In practice, winning parties do not bear the costs of losing parties, and the vast majority of arbitrators followed this convention. In effect, there are only three reasonable outcome choices of interest that are available to arbitrators: the losing party pays all, the losing party pays some, and each party pays its own costs. However, perhaps because of a mistaken click on the part of the arbitrator, 13 of our arbitrators indicated that they would have the winner reimburse the loser. While we could drop those individuals, doing so would risk inducing bias, as we would be controlling for a posttreatment variable—the survey response. To avoid posttreatment bias, we leave those arbitrators in the analysis (and treat them as choosing none of the three outcome options), preferring to err on the side of slightly more measurement error in the outcome (and therefore variance) than systematic bias.

In our analysis, we collapse the outcome variable into these three response categories. Merging the outcome categories this way has the important benefit of improving estimation efficiency by reducing the number of outcome categories considered and allowing us to pool results from vignettes in which the claimant won and those in which the respondent won. We therefore focus on estimating the average effect of being appointed by the winning versus the losing party, averaging over the separate claimant and respondent effects.4

The affiliation-bias hypothesis suggests that arbitrators who were told that they were appointed by the winning party will be more punitive toward the loser in terms of cost allocation than those who were told that they were appointed by the losing party. To analyze the effect of being appointed by the dispute’s winner versus the loser, we simply take the difference in means for each of the three response indicators of interest between arbitrators assigned to each treatment condition.5 Because treatment is randomized, these differences in means are unbiased estimates

4. While it would be possible to estimate separate effects for claimant compared with respondent appointees, we have insufficient power to meaningfully detect much of a difference between the two. To obtain greater precision, we find it more appropriate to estimate a single pooled winner versus loser effect given the constraints of the sample size.

5. With a trichotomous outcome variable, another approach commonly used by researchers is to assume a model for an underlying latent variable that is then observed in a coarsened form. These ordered dependent-variable models (such as ordinal logit or probit), however, would require us to make additional functional form assumptions relating the covariates to the outcome. Moreover, the proportional odds assumption for ordered logit models may not be satisfied in this case, as the effect of treatment is not constant across units. Our results suggest that arbitrators who would choose to split costs are unaffected by treatment, but those who would choose reimbursement are moved to select a
for the causal effect of the appointer treatment on the probability that an arbitrator would select a particular response.

Figure 1 presents the estimated treatment effects of winner versus loser appointment and party versus blind appointment on the three cost-allocation outcome categories. Thin lines denote 95 percent bootstrapped confidence intervals (5,000 iterations), and thick lines denote 90 percent bootstrapped confidence intervals. There were 72 arbitrators appointed by the winning party, 55 appointed by the losing party, 67 appointed jointly, and 63 appointed blind. On average, arbitrators were about 18 percentage points more likely to award all costs to the winning party when they were appointed by the winner rather than the loser. However, assignment to the winning party did not appear to change arbitrators’ propensity to cost shift in the first place. When an arbitrator chose to follow the unwritten rule of costs following the event, the winner’s appointee was more likely to punish the losing party by having it reimburse all of the winner’s costs, while the loser’s appointee was more punitive option. Because we use an experiment, any outcome modeling approaches (such as a logit model) are unnecessary at best and misleading at worst.

6. See Table B1 for a summary of the mean outcome responses by treatment condition. See Volfovsky, Airoldi, and Rubin (2015) for an extended discussion of defining treatment effects on ordered but discrete outcome variables.
likely to protect its appointing side by having it pay only some of the winner’s costs. This is consistent with the affiliation-bias story. While arbitrators do not completely advance their appointing party’s interests, when room for discretion arises, they appear to be more likely to choose outcomes that are more favorable to the side that appointed them.

It is worth noting that these results are likely a very conservative test of affiliation effects. Arbitrators taking the survey are not actually participating in a months-long proceeding that may reinforce party allegiances, nor do they face any potential costs or benefits to how they rule in the vignette. The bias we are able to detect is purely implicit and inherent in the role assignment. Moreover, if arbitrators are sensitive to the stigma of being perceived as biased, then they would likely try to attenuate any overt expressions of bias in their responses. If there is any form of social desirability bias influencing surveyed arbitrators, then it would likely bias the experiment against finding a difference between the treatment conditions. Nevertheless, we find strong evidence for a party-affiliation effect.

In addition, we wanted to consider whether blind appointees behave differently from those appointed jointly by the parties. As shown in Figure 1, while party appointees were slightly more likely to not shift costs, this difference is not statistically significant at any conventional rejection level. Overall, the magnitude of any difference between the two groups appears to be quite small, which suggests that blinded arbitrators are likely to behave much like the joint-party appointees when given discretion over some resource allocation between the parties. This suggests that blinding could have the benefit of having the parties agreeing to an arbitrator without facing the costs associated with such agreement.

One concern with the first experiment is that it may be difficult to generalize the observed behavior in response to a vignette on cost to real-world arbitral decisions on questions that implicate the substance of the case. Because fees are a comparatively less significant component of the decision relative to the merits and damages and arbitrators are much less constrained in how they choose to award costs, it may be that observed biases dissipate when arbitrators consider questions of greater importance to the case. While our argument for choosing to look at costs is precisely that the absence of constraints (that is, discretion) is most likely to generate affiliation bias and it is important to note that legal costs are not an insignificant component of the parties’ overall expenses in a dis-

7. We thank a reviewer for this point.
pute, we did consider it valuable to determine whether our result is robust to different choices of legal questions.

Therefore, we conducted a replication study in March 2017 that was designed to evaluate how arbitrators would reason through a decision on damages, that is, how much the claimant should be awarded given that the tribunal has found the respondent at fault. As in the first experiment, we chose to ask arbitrators to decide a discrete question to keep the vignette within a reasonable length and the task as clearly defined as possible, but this time using a question about damages. Asking arbitrators to evaluate the entirety of a case would still be unreasonable in the time and space allotted for the experiment. However, a decision on damages provides a tougher test of the affiliation-bias hypothesis. Arbitrators certainly face greater legal constraints and actual precedential guidance when evaluating how much a successful claimant should be awarded in compensation. Moreover, the amount of damages awarded to a claimant is certainly a central component of any dispute and highly salient to the litigating parties. Therefore, this vignette serves as an initial test of how well our affiliation-bias finding extends to more substantive decisions that compose an arbitration dispute.

Appendix A presents the vignette we used in the second experiment. In this vignette, a similar number of participants were presented with a setup similar to that in the first experiment. Arbitrators were told that the tribunal to which they were appointed had decided unanimously that the claimant-investor was entitled to damages because of the respondent state’s violation of the fair and equitable treatment standard of a hypothetical treaty. They were then tasked with deciding how much the claimant should be awarded in damages, one possibility proposed by the claimant and another proposed by the respondent. The key legal question that arbitrators had to evaluate was whether the claimant should be awarded compensation for lost future profits (the claimant’s argument) or whether the award should be based exclusively on the liquidation value of the firm (the respondent’s argument). According to the vignette, both parties cited relevant ICSID precedents for their positions. For a frame of reference, we included a link to the precedents and an article describing the typical award amounts in ICSID disputes. As with the first vignette, our goal was to provide enough information so that arbitrators would feel comfortable providing some answer but not so much information that there would be no variation in the arbitrators’ responses. We also retested our original question about costs.
Because our interest is in replication, in this vignette we manipulated only two elements: the appointing party and the amount of the proposed damages. As before, all treatments were randomly assigned independent of the surveyed arbitrators’ characteristics. However, to increase statistical power for detecting the affiliation effect, we chose to assign two-thirds of arbitrators to a party-appointed condition (rather than one-half, as in the first experiment). We included the manipulation on damages to evaluate whether differences in the magnitude separating each of the parties’ proposals affected the choice of damages or moderated the affiliation-bias effect. We wanted to ensure that our findings would not be driven exclusively by the values we chose for the claimant’s and respondent’s proposed damages. However, we found no evidence for effect modification or a statistically significant effect of the size of the proposed damages on the probability that the arbitrator would choose either the claimant’s or the respondent’s position. Table 2 summarizes the number of observations in each treatment condition.

After reading the vignette, arbitrators were asked to choose the amount that the respondent should pay in damages because of the treaty violation. We constrained our participants to choosing exclusively between the claimant’s proposal and the respondent’s proposal. In practice, investment arbitrators have the freedom to choose the amount awarded and rarely give the exact value proposed by either party. However, to increase the statistical power in our experiment, reduce variance in the out-

Table 2. Summary of Conditions in Experiment 2

<table>
<thead>
<tr>
<th>Variable and Conditions</th>
<th>Question on Damages (N)</th>
<th>Question on Costs (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appointer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appointed by the respondent</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td>Appointed by the claimant</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>Joint appointment</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Blind appointment</td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>Respondent’s proposed damages (US$):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25,001,050.00</td>
<td>85</td>
<td>86</td>
</tr>
<tr>
<td>12,500,525.00</td>
<td>70</td>
<td>73</td>
</tr>
<tr>
<td>6,250,262.50</td>
<td>93</td>
<td>93</td>
</tr>
</tbody>
</table>

**Note.** Chi-squared tests for the marginal counts across all four conditions fail to reject the null hypothesis that the counts are generated by discrete uniform distributions ($p > .10$). $N = 257$ arbitrators.
come, and increase the response rate by simplifying the task, we decided to limit the arbitrators to two choices. This also forced arbitrators to consider the arguments of the parties instead of simply taking the shortcut of splitting the difference between the two proposals. Since we were interested in seeing whether our original results on costs were replicated, we also asked our arbitrators to decide how the parties’ costs should be apportioned using the same choices of outcome as in the first experiment. We recruited our arbitrators using the same pool of e-mail addresses as in the first experiment, removing any arbitrators who unsubscribed from the list or notified us that they were not interested in participating in any surveys. A total of 644 arbitrators consented to participate. However, as before, not all arbitrators finished the survey. Only 248 of the arbitrators who began completed the survey and responded to the questions on damages, and 252 arbitrators answered the replication of the question about costs.

Figure 2 plots the estimated effect of being appointed by the claimant (the winner) versus the respondent (the loser) on the probability of choosing the claimant’s (high) proposed damages along with the estimated effect of party versus blind appointment. Again, thin lines denote 95 percent bootstrapped confidence intervals (5,000 iterations), and thick lines denote 90 percent bootstrapped confidence intervals. There were 88 arbitrators appointed by the respondent, 79 appointed by the claimant, 45 appointed jointly, and 36 appointed blind. On average, arbitrators appointed by the claimant were about 15 percentage points more likely to choose the claimant’s damages proposal compared with arbitrators appointed by the respondent ($p = .055$). While the result is just barely insignificant at the .05 level, we can reject the null hypothesis at just about any slightly higher rejection threshold.

The evidence from the follow-up survey is strongly suggestive that affiliation bias remains even for more substantive questions and beyond just decisions on costs. As before, there does not appear to be a sizeable difference between blind and joint-party appointees in their choice of damage awards. However, it may be the case that the effect is attenuated

8. A total of 25,965 recruitment e-mails were sent out in the second wave. As in the first experiment, many addresses were inactive or unavailable. Because we did not store identifying information from the first wave, we could not exclude arbitrators from the first experiment. However, we do not think that there are likely any interference effects between experiments given the year-and-a-half-long gap between them.

9. See Table B2 for a summary of the mean outcome responses by treatment condition.
somewhat by reduced discretion and greater precedential constraints associated with decisions on damages versus costs. Nevertheless, a promising result from our replication is that surveyed arbitrators’ answers to the costs question are nearly identical to those from our original study. In the replication, we exposed arbitrators only to a claimant-wins condition, while in the original study most outcomes were conditions in which the claimant failed to obtain a positive outcome. As shown in Figure 3, arbitrators appointed by the winner (claimant) were about 15 percentage points more likely to have the respondent reimburse all of the claimant’s costs. As before, thin lines denote 95 percent bootstrapped confidence intervals (5,000 iterations), and thick lines denote 90 percent bootstrapped confidence intervals. There were 90 arbitrators appointed by the respondent, 79 appointed by the claimant, 45 appointed jointly, and 38 appointed blind. Figure 3 shows that affiliation bias drives the magnitude of the cost award but does not affect arbitrators’ initial decision on whether costs should follow the event.

However, we also found in the replication experiment that blind appointees differed somewhat from joint appointees in how they apportioned costs. Joint appointees were significantly less likely to choose to have the respondent reimburse all of the claimant’s costs relative to blind appointees. While we saw weak evidence of this in the first experiment, the difference between the two groups is much larger in the replication. Yet, because the sample size allocated to the blind and joint-appointee conditions is smaller (since we wanted more power to detect the affiliation-bias effect), it is important not to read too much into this finding. In general, with smaller samples, we can expect more variable and potentially more extreme estimates. Nevertheless, this surprising result

Figure 2. Estimated average effects of appointment on the probability that respondents select the claimant’s damages proposal: experiment 2.
does suggest one possible caveat to the observed similarity between blind and joint appointees that we observed in the first experiment—when one party wins outright, joint-party appointees may have an aversion to extreme cost awards and favor a slightly more middle-of-the-road compromise relative to blind appointees.

We summarize the results of both experiments in Table 3. Overall, the evidence strongly suggests a meaningful affiliation effect when arbitrators are tasked with allocating some amount between the parties—a fundamental aspect of arbitration. When given strong discretion, as is the case for cost awards, party appointees tend to give the party that appointed them a more favorable outcome. Winning-party appointees demand more from the loser, while losing-party appointees try to mitigate their appointer’s losses.

3.2.3. External Validity. The experimental design allows us to isolate the treatment effect from selection biases that is common to observational analyses. Our estimates have high internal validity. We are confident, by design, that the observed association is not due to unobserved factors. Moreover, we obtained a large sample of a selective group of legal actors—to our knowledge the largest set of experiments conducted in this population. However, it is also important to assess the external validity
of our estimates—the extent to which our findings can be generalized to
the population of interest.

In the case of our experiment, we want to generalize how arbitrators
behaved in response to our vignette to the population of arbitrators serv-
ing on investor-state tribunals—the larger context of current blinding de-
bate and the experiment. If nonresponse to our survey was random, then
we could treat the observations as a random sample from the true popu-
lation, and our sample average treatment effect would be an unbiased es-
imate of the population average treatment effect. Even if nonresponse is
correlated with arbitrators’ characteristics, our experiment still yields an
unbiased estimate of the sample average treatment effect. This still cap-
tures the average effect of treatment in a theoretically interesting subpop-
ulation within the total set of arbitrators. While testing for nonrandom
nonresponse is impossible, we assess how representative our sample is
by comparing the background characteristics provided to us by surveyed
arbitrators with those in the population of interest, namely, ICSID arbi-
trators.

We consider three sets of covariates that we were able to measure for

Table 3. Summary of Treatment Effects of Party Appointments

<table>
<thead>
<tr>
<th>Experiment 1:</th>
<th>Award on Costs</th>
<th>Award on Damages:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winner versus loser</td>
<td>.019 (.080)</td>
<td>-.188* (.076)</td>
</tr>
<tr>
<td>Joint versus blind</td>
<td>.043 (.081)</td>
<td>-.030 (.076)</td>
</tr>
<tr>
<td>Experiment 2:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winner versus loser</td>
<td>-.050 (.066)</td>
<td>-.151* (.060)</td>
</tr>
<tr>
<td>Joint versus blind</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Point estimates denote differences in the share of arbitrators choosing each outcome. Bootstrapped standard errors are in parentheses (5,000 iterations).

a Claimant is the winning party; respondent is the losing party.

* p < .1.

+ p < .05.
most of our arbitrators: gender, their country’s legal tradition, and employment background. Of the 257 arbitrators from the first experiment, 240 provided information on these covariates. We then obtained a list of all 188 arbitrators who served on an ICSID tribunal constituted between 2010 and 2015 to act as our target population. For each of these variables, we calculated the proportion of arbitrators in each category and compared the target proportions with the proportions in our sample. Figure 4 compares the two covariate distributions. Crosses denote the true proportion among ICSID arbitrators. Lines denote 95 percent confidence intervals for our sample estimates.

Overall, the distribution of ICSID arbitrators in recent cases is predominantly male. Only 9.5 percent of arbitrators who served on at least one ICSID tribunal in the 2010–16 period were women. This skew is also evident in our sample, which is very close to the distribution in the population. However, our sample does contain a slightly larger share of female arbitration experts—about 14 percent. Nevertheless, it is unlikely that this small discrepancy is sufficient to give misleading estimates, particularly as we find no statistically significant difference ($p > .1$) in treatment effect magnitude between men and women.

Arbitrators’ nationality as a proxy for their country’s legal traditions is the second variable we considered. Consistent with Pauwelyn’s (2015) description of the distribution of arbitrators’ nationalities, the majority of arbitrators were nationals of European or North American countries. However, a fair number of arbitrators were from South American, Asian, or African countries. Our sample is not exclusively composed of nationals from a single state, which bodes well for the generalizability of our findings. One concern for our sample is that because the survey was administered in English, we may be more likely to get arbitrators from English-speaking countries and, in particular, countries with English common-law traditions. This could be an issue if arbitrators trained in different legal cultures approach the question of cost allocation differently. Indeed, the principle that costs follow the event or loser pays is well entrenched in some countries in the common-law tradition (see, for example, Woodroffe 1997).

Determining each arbitrator’s legal training is challenging given that nationality is not a perfect proxy—many arbitrators attend foreign law schools to obtain a master’s or doctorate degree abroad after a law degree. However, for the data that we were able to collect, it does appear to be a reasonable initial proxy. For each arbitrator who reported his or her
Figure 4. Characteristics of arbitrators ($N = 240$) in the sample: experiment 1
nationality or nationalities, we coded whether that country’s legal system has a common- or civil-law tradition using the data set of La Porta, Lopez-de-Silanes, and Shleifer (2008). We did the same for the 188 arbitrators in our target population. Among ICSID arbitrators, slightly more than half come from exclusively civil-law backgrounds, while about 41 percent had common-law backgrounds.

In our sample, these proportions are essentially reversed, which suggests that arbitrators with English common-law backgrounds were a bit more likely to respond to our survey. However, the magnitude of this difference is not particularly large. We also do not find evidence of effect heterogeneity for arbitrators with common- versus civil-law backgrounds. Testing for an interaction between our treatments does not yield a difference in effect that is statistically distinguishable from 0 at commonly used thresholds (p > .1). While this does not disprove the possibility of effect heterogeneity, it does suggest that our observed evidence for bias is unlikely to hold for only one major legal tradition. Because any effect heterogeneity is likely small or nonexistent, the difference between arbitrators’ legal origins in the sample and in the population does not meaningfully impact external validity.

Finally, we consider our arbitrators’ employment backgrounds and expertise. Respondents to our survey had four options to indicate their current area of employment: private sector, academia, government, or other. These categories were chosen to reflect the most common career backgrounds of international arbitrators as found by Fontoura Costa (2011). Arbitrators who answered “other” were recoded into one of the three remaining categories on the basis of their open-ended response. Most arbitrators answering “other” described themselves as independent arbitrators and were recoded as being in private practice. A few also noted employment in an international organization and were recoded as public-sector employees (as were those selecting “government”). We also coded the 188 ICSID arbitrators in our target population into one of the three career categories on the basis of their most recent area of employment, using publicly available information (for example, websites or curriculum vitae).

As shown in Figure 4, the majority of arbitrators work in the private sector. While this is also the case among ICSID arbitrators, a larger

10. For accuracy, we treated the coding of socialist legal systems as another form of a civil-law system.
share of the target population originates in academia or the public sector. About 92 percent of our arbitrators indicated private-sector employment compared with 69 percent in the ICSID group. While this does suggest a potential limitation of our sample, it is the case that arbitrators with backgrounds in the private sector make up the largest part of the overall international arbitration pool. Even if there is a difference in the response of private-sector arbitrators relative to those employed in other areas, our experiment credibly identifies the treatment effect for a highly meaningful subgroup of arbitrators—those employed in the private sector. Moreover, as with the other two factors considered, we are skeptical of a strong influence, as we do not find strong evidence that treatment effects differ between private-sector arbitrators and those not in the private sector. Therefore, despite the overrepresentation of private-sector arbitrators, our sample does not appear to be particularly idiosyncratic in a way that could give misleading results compared with the international arbitration population at large. We are confident that our experimental approach demonstrates that individuals are affected by affiliation bias in arbitration.

4. LIMITATIONS

Our evidence suggests that blinding could be an effective debiasing policy alternative to improve the practice of arbitration. Not only can blinding correct the observed affiliation effects that result from unilateral party appointments, but parties maintain the ability to make unilateral party appointments—one of the oft-cited advantages of arbitration compared with more conventional forms of litigation. Moreover, the practice of blinding can be easily extended to all members of an arbitration tribunal. By doing so, individual arbitrators may face significant uncertainty about the source of their appointment, having to consider not only the possibility of being a party appointee but also the potential of being a joint appointee or an appointee of a neutral authority.

Nevertheless, there are some clear limitations with our experimental approach in particular and with blinding as a practical solution more generally. Our experiment is unable to assess exactly how implicit biases of individual arbitrators affect outcomes of a body issuing a collective decision. Prior research on this question suggests that in deliberations, biases tend to be reinforced—the bias-accentuation effect (Fiedler and Kutzner
Hence, the implicit biases of the two party-appointed members may not simply cancel one another out, as often assumed, but rather may become further entrenched throughout the arbitration process.

There are also significant implementation challenges to blinding in practice as explained by Puig (2016). But, even if implemented successfully, blinding could only ameliorate affiliation effects. Hence, it remains to be seen whether a reform of this nature will substantially reduce the partisanship observable in many settings of arbitration. There are many reasons for that, but the main one is that selection effects would still persist with blind appointments. Litigants are aware of the preferences of specific arbitrators, especially of repeatedly appointed arbitrators, and appoint this core group precisely because they are more predictable and effective in signaling a particular position, by, for example, voting or dissenting. In fact, some arbitrators might prefer to signal their political preference clearly via their decisions or academic writings; hence, they are less willing to compromise and less reluctant to dissent when in disagreement with the majority.

Finally, to remain a part of a professional community, arbitrators rely heavily on individual reputation and social capital. Once members of the profession develop a reputation, this information is passed on by other members of the profession and often translates into other nominations precisely because of such leanings. Hence, while the thick social structure of the arbitral community may be valuable for facilitating the dissemination and development of new practices and professional norms, it may also create additional hurdles by accentuating the selection effects.

5. CONCLUSION

In the field of arbitration, the frequent use of party-appointed arbitrators is likely to result in litigant-induced biases. Disentangling selection effects (parties appointing friendly arbitrators) from affiliation effects (arbitrators changing their behavior in response to their appointment) is particularly challenging using observational data. While the former is inherent in the system of party appointment, the latter could be ameliorated by blinding arbitrators—restricting party appointees from knowing the source of their appointment throughout the proceedings.

Through novel survey experiments, we provided a methodological solution to the problem of measuring affiliation effects in a world confounded by selection effects. As our results show, being appointed by one
of the parties in a dispute directly changes the behavior of arbitrators. Hence, the appointment itself is the cause of some of the bias toward the arbitrator’s appointing party. Apparent patterns of bias in real decisions are unlikely to be simply an effect of parties filtering their choice of appointee on the basis of known, prior attitudes.

Normatively, our results provide support for implementing blinding in arbitration proceedings. While the implementation challenges do not seem insuperable, the effectiveness of blinding may also be heavily dependent on the social and institutional context of arbitration. Nevertheless, our article should be a precursor for assessing affiliation bias and the impact of blind appointments in more realistic settings. The proposed methodology—we believe—is a significant contribution to exploring questions of judicial politics and the psychology of decision makers’ biases.

**APPENDIX A: SAMPLE EXPERIMENTAL VIGNETTES**

**Experiment 1**

Imagine an investor-state dispute being conducted under the 2006 Arbitration Rules of the International Centre for Settlement of Investment Disputes (ICSID). The Claimant is a firm headquartered in a high-income economy. The Respondent is a country classified by the World Bank as a middle-income economy.

The Claimant alleged that the Respondent violated the provisions of a bilateral investment treaty to which the Respondent is a party. Among other arguments, the Claimant argued that the investor and its investments had been treated unfairly and that ultimately the Respondent expropriated the Claimant’s investment located within the Respondent’s territory. The underlying dispute concerns an infrastructure project undertaken by the Claimant under a concession contract with a governmental agency. The Respondent argued in response that the Claimant had violated provisions of the contract and that the investors received all compensation to which they were entitled.

You were appointed to the Tribunal [by the Respondent.][by the Claimant.][by the Parties.][.]. After careful consideration of the facts of the case, the tribunal unanimously decided that the Respondent unfairly treated and wrongfully expropriated the Claimant’s investment and that the Claimant is entitled to compensation.

In their submissions on costs, both parties have requested that the other party bear the costs of the proceedings in full, including legal fees and expenses. The counsels for both parties behaved professionally and ethically during the proceedings.
Experiment 2

Imagine an investor-state dispute being conducted under the 2006 Arbitration Rules of the International Centre for Settlement of Investment Disputes (ICSID). The Claimant investor alleged that the Respondent state violated the provisions of a bilateral investment treaty to which the Respondent is a party. Among other arguments, the Claimant argued that the Respondent violated the treaty’s fair and equitable treatment provisions and mistreated the Claimant’s investment. You were appointed to the Tribunal [by the Respondent.]/[by the Claimant.]/[by the Parties.]/[.] After careful consideration of the facts of the case, the tribunal (you and your fellow arbitrators) unanimously decided that the Respondent unfairly treated the Claimant’s investment in violation of the treaty and that the Claimant is entitled to compensation.

You are now asked to decide on the amount of damages owed to the Claimant by the Respondent. The parties have agreed that the tribunal’s task is simply to pick one of the two positions of the parties’ experts and decide how the expenses should be apportioned in this dispute. In its relevant part, the bilateral investment treaty provides as follows:

1. A Tribunal may award monetary damages and any applicable interest, only.
2. A Tribunal may also award costs in accordance with the applicable arbitration rules.
3. A Tribunal may not order a Party to pay punitive damages.

The Claimant has argued that they should be compensated for lost future profits that would have been realized had the measure not taken place plus interest. The Claimant justifies this claim on the grounds that the enterprise operated profitably for a period of almost three years prior to the violation. The Claimant cites Metalclad v. Mexico, ICSID Case No. ARB(AF)/97/1, which considered a minimum presence of at least two or three years necessary for an award of future profits. The Claimant’s expert has calculated damages for US$50,002,100.00 based on the discounted cash flow value of the expected returns from the Claimant firm’s ten-year investment plan. The Respondent has argued that the enterprise had not operated for a sufficient period of time to establish itself as a “going concern” and that the ability of the enterprise to generate future earnings was uncertain and compromised. The Respondent cites Tecmed v. Mexico, ICSID Case No. ARB(AF)/00/2, arguing that the tribunal in that dispute ruled that the Claimant’s operating history of two and a half years was insufficient to establish enough objective data on profitability to apply a discounted cash flow analysis. Therefore, any estimate of future profits would be highly speculative. The Respondent instead proposes that damages should be based on the liquidation value of the firm and their expert has calculated damages for [US$25,001,050.00]/[US$12,500,525.00]/[US$6,250,262.50].
In ICSID disputes, the average award for Claimants who are awarded damages is about US$45.6 million. The median award is US$10.9 million (See: Franck, S. D., & Wylie, L. E. (2015). Predicting outcomes in investment treaty arbitration. Duke Law Journal, 65(3), 494–527). Throughout the proceedings, both disputing parties were cooperative and the counsels for both parties behaved efficiently, professionally and ethically. The parties have not agreed on how and by whom the expenses shall be paid.

APPENDIX B: ADDITIONAL TABLES

Table B1. Share of Respondents Choosing Each Outcome by Treatment Condition:
Experiment 1

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Appointed by Winner (n = 72)</th>
<th>Appointed by Loser (n = 55)</th>
<th>Joint Appointment (n = 67)</th>
<th>Blind Appointment (n = 63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Split costs</td>
<td>.292 (.053)</td>
<td>.273 (.061)</td>
<td>.328 (.058)</td>
<td>.286 (.057)</td>
</tr>
<tr>
<td>Loser pays some costs</td>
<td>.139 (.041)</td>
<td>.327 (.064)</td>
<td>.224 (.051)</td>
<td>.254 (.055)</td>
</tr>
<tr>
<td>Loser pays all costs</td>
<td>.542 (.059)</td>
<td>.364 (.066)</td>
<td>.358 (.060)</td>
<td>.413 (.062)</td>
</tr>
</tbody>
</table>

Note. Bootstrapped standard errors are in parentheses (5,000 iterations).

Table B2. Share of Respondents Choosing Each Outcome by Treatment Condition:
Experiment 2

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Appointed by Winner (Claimant) (n = 79)</th>
<th>Appointed by Loser (Respondent) (n = 90)</th>
<th>Joint Appointment (n = 45)</th>
<th>Blind Appointment (n = 38)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claimant’s damages proposal</td>
<td>.544 (.055)</td>
<td>.398 (.052)</td>
<td>.489 (.074)</td>
<td>.500 (.085)</td>
</tr>
<tr>
<td>Split costs</td>
<td>.228 (.047)</td>
<td>.278 (.049)</td>
<td>.222 (.062)</td>
<td>.286 (.057)</td>
</tr>
<tr>
<td>Loser pays some costs</td>
<td>.127 (.038)</td>
<td>.278 (.047)</td>
<td>.333 (.071)</td>
<td>.254 (.055)</td>
</tr>
<tr>
<td>Loser pays all costs</td>
<td>.557 (.055)</td>
<td>.400 (.051)</td>
<td>.333 (.071)</td>
<td>.413 (.062)</td>
</tr>
</tbody>
</table>

Note. Bootstrapped standard errors are in parentheses (5,000 iterations).
REFERENCES


