INDIVIDUAL VS. GROUP DECISION-MAKING: EVIDENCE FROM A NATURAL EXPERIMENT IN ARBITRATION PROCEEDINGS

Naomi Gershoni

Discussion Paper No. 19-12

November 2019

Monaster Center for
Economic Research
Ben-Gurion University of the Negev
P.O. Box 653
Beer Sheva, Israel

Fax: 972-8-6472941
Tel: 972-8-6472286

Individual vs. Group Decision-Making: Evidence from a Natural Experiment in Arbitration Proceedings

Naomi Gershoni*

November 11, 2019

Abstract

The importance of understanding the systematic differences between group and individual decisions has been well recognized in the literature. However, the vast majority of empirical evidence on this issue is derived from laboratory experiments, and hence does not reflect professional incentives and career concerns, both of which may play a crucial role. To fill this gap, I exploit a unique regulatory change that exogenously decreased the number of presiding arbitrators from three to one for a specific class of cases in the Financial Industry Regulatory Authority arbitration as well as an original data set of arbitration awards. The findings indicate that sole arbitrators tend to render more moderate awards when compared to panels of three arbitrators. Adding arbitrator fixed-effects to the model confirms that this tendency is also present *within* arbitrators, implying that the same arbitrators are inclined towards more extreme "all or nothing" decisions when in groups. This finding rules out the possibility that the effect is driven by differences in the selection of arbitrators into panels. Rather, evidence supports a novel explanation of the polarization of groups: namely, individuals' concerns about adverse effects of extreme decisions on their reputation are mitigated within groups, in which individual opinions are at least partially obscured.

*Ben-Gurion University of the Negev, naomige@bgu.ac.il. I am grateful for the advice of Zvika Neeman and Analia Schlosser. I also thank Alon Klement, Danny Cohen-Zada, Shirlee Lichtman-Sadot and the participants of SOLE 2018 and ALEA 2019 for helpful comments. I gratefully acknowledge financial support from the Israeli Science Foundation (grant no. 1394/17).

1 Introduction

Understanding the systematic differences between group and individual decisions has been of interest in many disciplines for several decades now. Economic decisions are made by groups in a wide array of settings, including business decisions by boards, policy recommendations by committees, and household choices. While common wisdom suggests that "two heads are better than one," the theoretical and experimental literature has long acknowledged that group decisions are not necessarily superior and that more intricate psychological and economic mechanisms may drive differences between group and individual decisions. Moreover, such differences often depend on context, on institutional design, and on the criteria by which the decision is evaluated.[1] In light of the numerous competing predictions, the scarcity of empirical evidence based on "real world" observational data is striking. However, this is not surprising since the formation of groups, as well as the type of decisions they make, are typically endogenous and hence evidence often points toward correlations rather than causal effects. At the same time, empirical evidence derived from laboratory experiments fail to reflect professional incentives and career concerns, which may play a crucial role.

This paper aims to identify the causal effects of groups on outcomes in the context of arbitrators' decisions, by exploiting a unique regulatory change in the Financial Industry Regulatory Authority (FINRA) arbitration that exogenously decreased the number of arbitrators assigned to a specific group of cases from three to one. The study analyses an original data set of arbitration awards in cases that involve investment firms and their customers.

The focus is on whether panels of three tend to be more or less extreme in their decisions when compared to sole arbitrators. Therefore, first and foremost, this study contributes to the body of evidence and long-standing discussions of the *group polarization* phenomena, which date back to Stoner (1961). Numerous lab experiments document that groups are likely to polarize the opinions of their individual members and demonstrate two potential *choice shifts* of groups, either "risky" or "cautious", depending on group members' individual predispositions and on the extent of deliberations (see e.g. Stoner, 1968; Fraser et al., 1971; Moscovici and Zavalloni, 1969; Kerr et al., 1996). However, Adams and Ferreira (2010), the only study I am aware of that systematically compares individual and small-group decisions in the field report a *moderating* effect of groups on outcomes. And when it comes to judicial decisions, studies of mock-juries remain largely inconclusive (for a discussion of these findings see Zamir and Teichman, 2018).

The debate about group polarization is particularly pertinent in the context of arbitrators. Sem-

[1] For some examples see Charness and Sutter (2012); Goeree and Yariv (2011); Eliaz et al. (2006); Cooper and Kagel (2005).

inal studies of arbitrators' behavior focused on the notion that arbitrators may or may not simply choose to "split the difference" between disputing parties, regardless of the evidence presented to them (Farber, 1981; Farber and Bazerman, 1986; Bazerman and Farber, 1985). Moreover, due to information asymmetry, extreme decisions may be avoided even when accurate because such decisions damage arbitrators' reputation as fair and unbiased (Posner, 2005; Klement and Neeman, 2013).[2] This raises the question of whether these inclinations depend on a decision being taken by an individual arbitrator or by a group.

FINRA arbitration is an ideal setting to study this question, first, because the number of arbitrators is determined according to a sharp regulatory cutoff, which is based on the amount of requested compensation. Second, the reputation of arbitrators is important because litigants choose their arbitrators and have the right to veto-out specific arbitrators (Egan et al., 2018). Combined with the fact that FINRA "customer cases" repeatedly confront two distinct groups with opposing interests — investors and brokerage firms, this system creates substantial incentives for arbitrators to maintain a reputation of impartiality by granting moderate "split the difference" awards.

In March 2009, FINRA raised the threshold for the assignment of cases to three arbitrators rather than one from \$50,000 to \$100,000. This change in regulations serves as a "natural experiment" that allows me to test the impact of the number of arbitrators on outcomes for cases between the two thresholds which were assigned to panels of three arbitrators before the change and to a sole arbitrator afterwards. An important feature of this research design is that the change in regulations and its timing were not intended to affect arbitration results. Rather, the change was induced by an unusual load of cases during the year 2009, which probably followed the financial crisis that peaked in the preceding year. Although this allows the rule change to be considered as an exogenous shock to tribunal size, it raises concerns that any observed pattern will be a result of events provoked by the financial crisis. More generally, other factors may have caused a shift over time, or it could be that awards simply exhibit a time trend.

To address these concerns and confounding factors, I use a *difference-in-differences* framework.[3] Cases that were not affected by the change, namely those with a requested amount either below the old threshold or above the new one serve as natural comparison groups (either combined or alternatively). To account for the possibility that different arbitrators are selected to serve as sole arbitrators and as panel members, I introduce arbitrator fixed-effects. Multiple robustness and placebo tests are conducted to confirm that the treatment and comparison group would be

[2] Morris (2001) and Ely and Välimäki (2003) predict "adverse reputation effects" when only the actions of agents are observed and the truth remains unknown. Also see Prat (2005) for a similar distinction between increased transparency of agents' actions and of their consequences.

[3] It should be noted that a regression discontinuity approach that compares cases just below and above the thresholds cannot be implemented efficiently in this setting, due to the small sample size.

expected to follow a similar trend if it were not for the change in regulations and that results are not driven by the events of the financial crisis. In addition, I confirm that there was no manipulation of the rules that determine treatment status and that the composition of the groups did not change differentially over time.

The main results clearly and robustly indicate that when compared to sole arbitrators, panels of three are approximately 20 percentage points more likely to deliver an extreme award. This increase is almost evenly split between extreme customer-wins that award claimants with the full amount of damages that they claimed for, and extreme losses that entirely dismiss their claims. However, because customer wins are relatively rare the difference that panels make for them is close to 100% whereas for losses, the difference is only 17%. Impacts of a similar magnitude are also found *within* arbitrators, indicating that differential selection into panels cannot account for the tendency of panels to polarize decisions. In addition, results hold when cases that were filed during 2009 following the financial crisis are excluded from the sample, indicating that the effects cannot be attributed to this event.

Various mechanisms have been suggested to explain differences between group and individual decision making. Adams and Ferreira (2010) suggest that group moderation could either be the result of a "compromise effect," since groups are expected to average-out extreme individual opinions in order to reach consensus, or of a "membership effect," which occurs when individuals with extreme opinions are excluded from groups (or of both effects). If these effects operate in the FINRA setting, according to the findings, they are obviously overridden by other channels that drive groups to polarize. When groups were found to polarize, the literature suggested various cognitive biases as a mechanism. One prominent example for such biases is the tendency of individuals to disregard the fact that information sources are mutual, a bias known as "correlation neglect" (Glaeser and Sunstein, 2009). Another well-known example is "social comparison," which is expected to promote conformity in groups.[4] However, groups also tend to mitigate common behavioral biases, as they decrease individual levels of social considerations and responsibility (Charness and Sutter, 2012). A case in point is from Arlen and Tontrup (2015), who find that groups overcome the "endowment effect" because shared responsibility decreases concerns for future regret. Shared responsibility may also drive groups of arbitrators to make more accurate extreme decisions that individual arbitrators might refrain from for fear of future regret.

I propose a novel mechanism for group polarization — group-reputation — and show that it is the most probable explanation to the findings of this study. The group-reputation channel could

[4] A more extreme version of this bias was first described by Janis (1982), who coined the term "groupthink" to describe the tendency of groups to apply self-deception in an attempt to conform with each other and avoid conflict. Bénabou (2013) presents a formal theoretical model of "collective denial and willful blindness in groups, organizations, and markets" that is in line with this idea.

be thought of as a variant of the mechanism suggested by Arlen and Tontrup (2015), except that it builds on concerns for reputation rather than regret and on perfectly rational considerations rather than behavioral biases. This channel arises in an environment where extreme decisions are always interpreted as (noisy) signals of bias because perpetual information asymmetry prohibits inference on the accuracy of decisions (or on their fairness). Hence, individual arbitrators try to avoid extreme decisions (even when they believe them to be correct) for fear of gaining a "bad" reputation as unfair or biased among potential litigants. However, the same arbitrators, as members of panels with a *secretive* deliberation process, can effectively "hide" behind the group and worry less about the reputational consequences of their extreme opinions. This could be because it is more difficult to figure out what each panel members' opinion actually was or because having three arbitrators agree on the outcome reduces suspicion for possible dishonesty or distortion. Therefore, extreme opinions will tend to be more pronounced in panels.

It has already been shown that reputation concerns in groups may lead to flows in information aggregation which under some circumstances result in polarization, or "herding" (Ottaviani and Sørensen, 2001; Da and Huang, 2018). In addition, it has been claimed that transparency is expected to increase such concerns (Levy, 2007) and was found to have an impact on levels of conformity (Meade and Stasavage, 2008). In regard to these studies, the novelty of the mechanism suggested here, is that individuals do not alter their stated opinions according to the opinions of other individuals in the group. Rather, they use the obscurity of the group to avoid a penalty that they would otherwise incur for expressing their genuine extreme opinion that could be either anchored in objective evidence or in a biased subjective view, regardless of other group members' stated opinions.

To establish that the group-reputation channel drives the main results, I distinguish between environments with and without reputation concerns, and show that group polarization only occurs when such concerns prevail. FINRA "industry cases" are (by definition) cases that do not involve customers and hence, both sides of the dispute are usually firms. Claimants (or respondents) in such cases do not share distinct qualities and cannot be identified as a part of either group before their claim was filed. Hence, no suspicion of favoritism towards any specific group can be raised. In this "reputation free" setting, no significant differences are detected between group and individual decisions.

While the group-reputation mechanism is identified with relative certainty, it remains unclear whether groups improve or worsen decisions in terms of accuracy (or fairness). On the one hand, biased arbitrators can "hide behind the group" and express their favoritism, hence reducing accuracy. However, when extreme decisions are warranted impartial arbitrators will not attenuate their views thus increasing accuracy.

The paper proceeds as follows. The next section describes the FINRA Arbitration setting and the unique data set. Section 3 presents the empirical strategy. Results, robustness tests, and evidence on mechanisms are discussed in section 4. Section 5 concludes.

2 FINRA Arbitration Data

2.1 Background

FINRA is a private corporation that acts as a self-regulatory organization for the financial industry by authorization of the Congress. The members of the corporation are securities firms and brokerages that operate in the US. In addition to other functions, FINRA runs the largest dispute resolution forum in the US securities industry, offering extensive arbitration services for disputes between its members and their customers (customer cases) or other disputes that involve member firms (industry cases). FINRA customer cases are considered to be mandatory arbitration proceedings, since customers are obligated to file their claims with FINRA according to standard investment agreements.[5]

According to FINRA regulations, claims above a specific threshold amount are assigned to three arbitrators whereas claims that are equal or below this threshold are assigned to one arbitrator. Parties may agree to stray from this rule, but in practice they rarely do so. The arbitrators are selected by the parties to the dispute using a "veto-rank" method, in which each party receives a list of ten randomly selected arbitrators, vetos up to four, and ranks the remaining six. Importantly, when three arbitrators are required, three different lists of ten arbitrators each are provided and the same process is applied for each list separately. Therefore, the probability for each party to get their favorite arbitrators appointed does not depend on the required number of arbitrators.

In addition, the regulations determine the types of arbitrators that should be assigned to the case. Each arbitrator in the FINRA roster is categorized as either "Public" or "Non-Public" according to the extent of his experience and involvement in the financial industry (e.g. as employee or attorney of a brokerage firm).[6] Only public arbitrators that meet certain requirements of experience and education can serve as chairpersons or as sole arbitrators.[7] In cases with three arbitrators, at least two arbitrators are public arbitrators and at least one must be eligible for chairperson.[8]

[5]see *Shearson/American Express Inc. v. McMahon* (482 US 220 [1987]), *Rodriguez de Quijas v. Shearson/American Express, Inc.* (490 US 477 [1989]).

[6]See rules 12100(p) & (u) of FINRA's Code of Arbitration Procedure for Customer Disputes.

[7]See rule 12400(c) of FINRA's Code of Arbitration Procedure for Customer Disputes.

[8]In most cases the panel of three arbitrators is a "Majority Public Panel", i.e. it consists of two public arbitrators and one non-public. Following a pilot program that started in 2008, in 2011 FINRA added the option to choose an "All Public Panel" namely three public arbitrators.

Panels decide according to a simple majority rule[9] after a discussion led by the panel chair. The FINRA guidelines require that each of the arbitrators in the panel "express their individual views of the case" and "take part in deliberating the facts and issues." Although unanimity is not required, the number of cases where a dissent is actually recorded is negligible. This is in line with the findings of previous studies that document and rationalize a tendency towards unanimity among members of judicial panels, a phenomena referred to as "dissent aversion" among judges (Posner, 2010; Epstein et al., 2011).

2.2 Data and Sample

The data was collected from FINRA arbitration awards, which are publicly available on-line. The database consists of awards in customer and industry cases filed between 2006 and 2011 and resolved by January 2013. All cases involve a claim for a specific amount of monetary compensation.[10] In addition to the number of arbitrators and their identities, each record includes filing and award dates, the relief requested in the statement of claim, and the awarded damages.

The main sample is restricted to customer cases with claims of up to $250,000.[11] For this sub-sample, detailed information on parties and cases was recorded, including the number of claimants and respondents and their representation, the causes of the claim (or controversy type), and the types of securities that are the subject of the dispute.

Awards are measured as rates, i.e., the fraction of the monetary relief initially requested by the claimants that was awarded by the arbitrators in their final award (hereinafter referred to as award rate or AR). Accordingly, the minimal award rate is zero, while the maximum is set to one, although it could potentially be higher if, for example, the arbitrators ordered respondents to pay punitive damages. In the interest of consistency, because the requested amount of punitive damages is rarely specified, the requested relief includes only compensatory damages. This definition is useful, since this is also the amount that FINRA uses to determine the number of arbitrators that are assigned to the case.

Award rates are used to define the main outcomes of interest. An extreme award is defined as an award rate that equals zero or one, meaning that one side to the dispute receives precisely what was requested. A moderate award is an award rate between 0.4 and 0.6 *or* a case that was resolved in settlement. Settlements are classified as moderate ARs, since a settlement can be interpreted as

[9] Rule 12410 of FINRA's Code of Arbitration Procedure for Customer Disputes.

[10] Cases in which the requested damages were not specified or where no monetary relief is requested are excluded from the analysis.

[11] The distribution of the requested damages has a very long right tail. Approximately 40% of cases have a relief requested that is higher than $250k with the maximal relief requested for cases in the full sample being 1.25 billion dollars. Hence, the 250k upper bound makes cases in the treatment and comparison groups more comparable.

each side receiving approximately half of what was requested.

There are two types of settlements in FINRA arbitration — ones that are announced in the written award and ones that are not recorded anywhere and are usually reached before the hearing stage. The first type is for cases in which parties settled at or after the hearing stage. Obviously, in such cases, arbitrators are expected to have an impact on the probability of settlement and thus, omitting these cases from the sample will bias the estimates. However, the details of the settlement are usually kept secret and therefore the award rate cannot be determined. When an exact award rate is required, these claims are coded as having award rates that equal 0.5, which is the average award rate for claims that were not entirely dismissed and did not resolve in settlement. As a robustness test, in section 4.2, I use observable characteristics to predict the AR for these claims, relying on the theoretical prediction that settlements occur when the outcome can be expected with relative certainty (Priest and Klein, 1984).

The second type of settlements is missing from the data due to its retrospective nature (in the sense that one only observes a dispute when the award is posted) and thus is not included in the analysis. More generally, approximately 80% of claims are closed before the hearing stage for reasons such as direct settlement, bankruptcy of a critical party, or forum denying, and therefore are absent from the data. Assuming that the selection of such cases is not correlated with the number of arbitrators, this does not pose a major threat (especially since the effect of arbitrators on outcomes at this stage is negligible). Appendix Figure A2 shows that the number of cases in the data is proportional to the total number of claims that were filed each year according to FINRA's aggregate statistics.

Data censoring is another important issue that arises due to the retrospective nature of the data. To avoid such a distortion, I use survival analysis and only consider cases for which the award was issued within a limited period of time since filing which is calculated as the 90th percentile of the distribution of case duration. Additionally, I only include filing dates that preceded the last award date in the sample (January 31, 2013) by at least this time period. To understand how this limitation may affect results, a sensitivity analysis was conducted and boundaries were derived for the main estimates (see section 4.2). In addition, the results were confirmed using both a sample without any restriction and with a more restricted sample that sets the limit at the 75th percentile.

Figure 1 presents the cumulative distribution of duration for three categories of cases. These categories are based on FINRA regulations regarding the number of arbitrators and correspond to the choice of treatment and control groups, as explained in the following section. To estimate the complete uncensored distributions, only cases that were filed during 2006 and 2007 are included, assuming that for these two years, 100% of cases that eventually resulted in a written award are observed in the sample. More specifically, this assumption is that the number of cases that take longer than five years to complete is practically zero.[12] The vertical line in each figure marks the 90th percentile, indicating that for the 50-100K group (which will serve as the treatment group), 90% of cases ended within 744 days. For the first comparison group, which is characterized by lower claims, the duration of proceedings tended to be shorter, and the opposite is true for the second comparison group.

3 Empirical Design

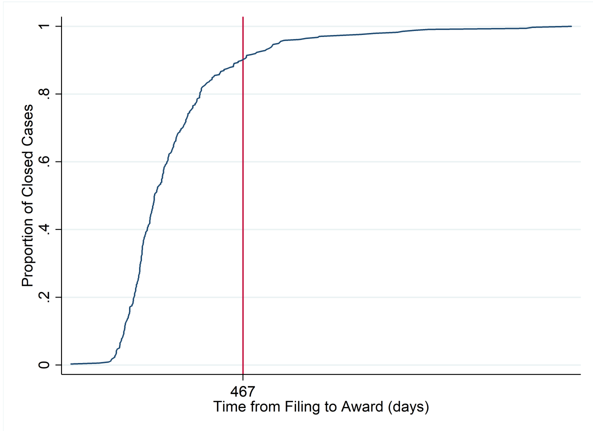
Up to March 29th, 2009 the threshold that was set by FINRA for a sole arbitrator was $50k and cases above this amount were assigned to panels of three. Figure 2 presents a 'naive' comparison of the distribution of awards granted by panels of three versus sole arbitrators, for this period, which in the main analysis will serve as the pre-change period. According to Figure 2, cases that were heard by sole arbitrators seem to have slightly more extreme awards compared to cases with three arbitrators. However, this difference cannot be directly attributed to the different number of arbitrators, since there are many other observed and unobserved qualities which distinguish them. Obviously, cases with one and three arbitrators differ in the amount of requested damages and this amount is expected to be correlated with the cause of claim, the type of investment, and with the characteristics of the investor and the firm (e.g. investor's income and wealth, firm's reputation or professional level).

In 2009, FINRA raised the threshold for the assignment of three arbitrators to $100k.[13] Press coverage and investment bloggers suggested that the change was required due to an unusual load of new claims, which was probably the result of the financial crisis. Indeed, in 2009 an unusually large number of claims were filed as can be observed both in FINRA aggregate statistics and in my sample (see Appendix Figure A2). This threshold change offers a unique opportunity to compare pre- and post-change arbitration awards in cases between the old and the new threshold, i.e. with a requested relief between $50k and $100k. For this group of cases, if a claim was filed until March 29, 2009, the rules assigned a panel of three arbitrators, whereas after this date, only one arbitrator was required.

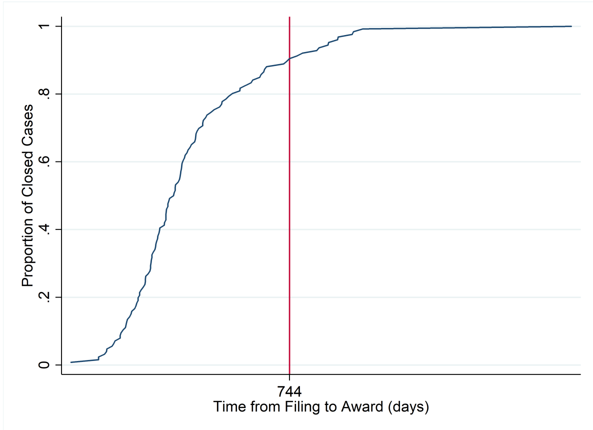
Figure 3 presents the distribution of award rates before and after the change in regulations for this group. It can be seen that there was a sizable shift in the distribution of awards and that the difference between decisions of panels and sole arbitrators is actually opposite to that suggested

[12]The awards in the sample were collected until January 2013, allowing a maximal duration of seven years for cases filed in the beginning of 2006, and five years to those filed very close to the end of 2007.

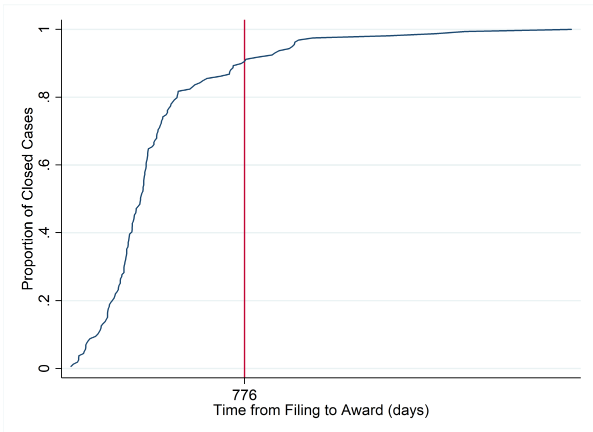
[13]See FINRA regulatory notice 09-13 "Threshold for Single Arbitrator Cases."



(a) $0 < RR \leq 50k$



(b) $50k < RR \leq 100k$



(c) $100k < RR \leq 250k$

Figure 1
Case duration CDF (by range of relief requested)

Notes: The figure presents the cumulative distribution function of the duration of proceedings (from filing to award). The distribution is presented separately for three categories of cases divided by the relief requested by the claimant. The vertical line in each figure marks the 90th percentile.

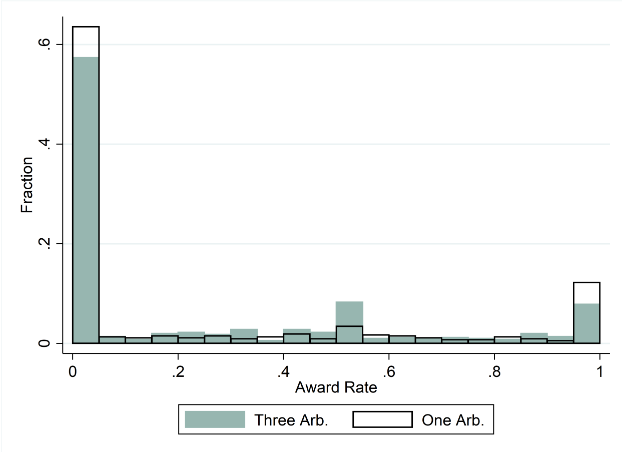


Figure 2
Distribution of AR by number of arbitrators, pre-change

Notes: The figure presents the distribution of award rates for claims that were filed between January 1, 2006 and March 29, 2009, by the number of arbitrators (three or one).

by Figure 2 (where selection is not controlled for). In the post-change period, when sole arbitrators were assigned to cases instead of panels of three, the fraction of cases for which claims were denied entirely (or with an award rate close to zero) substantially decreased. A similar decrease is observed for awards that grant claimants the full amount of damages they initially requested (or more). At the same time, awards that tend to split the difference, i.e., an award rate close to 0.5, become more frequent in the post-period.

While this finding can be explained by the change in the number of arbitrators, it can also be the result of some general time trend or some other unobserved heterogeneity of the pre- and post-periods. To refute this concern, I use two very natural comparison groups: cases with a requested relief below \$50k which are assigned to sole arbitrators throughout the entire period and those with a requested relief above \$100k which are assigned to panels of three arbitrators throughout the entire period. The two comparison groups are used in a difference-in-differences (DD) setting. In the main specification, in order to increase power and precision, these two groups are used together, while controlling for each category of cases separately to allow for variation in the average award rate across groups. In addition, as a robustness test, the same equation is estimated using each control group *separately*. If similar effects are found both when the treated cases are compared to cases with higher claims and with lower claims, it clearly implies that the change is not directly related to the size of the claim but rather to the change in the number of

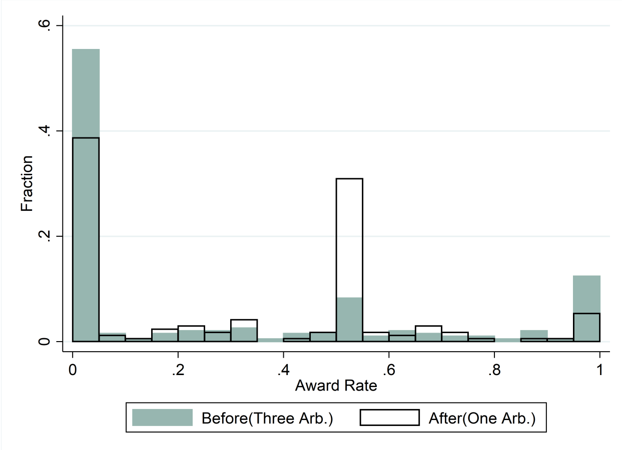


Figure 3
Distribution of award rate pre- and post-change for treatment group (all years)

Notes: The figure presents the distribution of award rates for claims in which the relief requested was between \$50k and \$100k. The claims were filed between 2006 and 2011. The distribution is plotted separately for the pre- and post-change periods (before and after March 30, 2009).

arbitrators. Therefore, the main specification for estimation is:

$$Y_{iq} = \alpha + \beta_0 Treated_{iq} + \beta_1 Post_{iq} + \beta_2 Treated \times Post_{iq} + \gamma' X_{iq} + \delta_q + \psi_{iq} \tag{1}$$

where Y is an indicator for an extreme (or moderate) award in case i which was filed during quarter q.

Treated is an indicator that the case is in the affected range of relief requested (i.e., \$50k to \$100k) and *Post* indicates whether the case was filed after the change of rules came into effect. $Treated \times Post$ is the interaction between the two indicators and is the explanatory variable of interest. Therefore, β_2 is the difference in outcome that is attributed to the assignment of a case to a sole arbitrator rather than to a panel of three arbitrators.

X is a vector of controls that can be divided into two groups. The first group consists of variables that refer to procedural and technical aspects: the number of claimants, parties' representation by an attorney, and whether or not a counter claim or a third-party claim was filed. The second group includes a large set of indicators for the type of controversy (cause of claim) and for the type of securities involved in the dispute. A complete list of these variables (and their mean values) is presented in Table 1. In addition, each specific category of cases within the comparison group is controlled for, including an indicator for cases with a relief requested equal to or below \$25k, for which the "simplified arbitration" procedure applies. Under this procedure, in order to accelerate the proceedings, claimants can choose to only submit written pleadings and evidence or to have a hearing conducted by a telephone conference call.[14] δ_q is a vector of filing-quarter fixed-effects that flexibly control for events that similarly impact all categories of cases over time.

Using data on arbitrators' identity, the same specification is then estimated with arbitrator fixed-effects. This specification addresses confounding factors that relate to unobserved arbitrator traits as well as to the potentially differential selection of arbitrators to serve as sole arbitrators (compared to panels). The estimated DD coefficient in this specification is the *within* arbitrator impact of deciding individually rather than in a group. It is important to note that the fixed-effects for panels are specified according to the identity of the panel chair, both because the same qualifications are required to serve as chair and as a sole arbitrator and because it is reasonable to assume that the chair has the greatest influence on the panel's decision (since she controls how the hearings and deliberation are conducted).

The main identifying assumption in the DD setting is that award rates for the treatment and comparison groups would follow a similar trend if the regulatory change had not occurred. This assumption seems plausible since treatment status is determined by somewhat arbitrary cutoffs set by FINRA regulations. While it is obvious that the rationale for the rules that govern the number of arbitrators is that higher stakes call for more decision-makers, there is no justification for the specific amount that dictates the sharp increase from one to three arbitrators. In addition, as noted above, the regulatory change was driven by the desire to alleviate caseload and therefore could be considered as random with regard to the outcomes of interest. It should be noted that the quazi-arbitrary cutoff calls for a regression discontinuity approach to identify causal effects. Unfortunately, the sample of cases close to the relevant cutoffs is too small to implement this approach efficiently.

The parallel trends assumption is tested by comparing pre-period trends between treatment and control groups, using the same DD specification only with the pre-period sample and a placebo treatment date. The results of this estimation are presented in Table 5 and clearly confirm the absence of differential pre-trends.

Although pre-trends are similar, when identification of treatment effects relies only on time variation, one should consider the possibility that other changes occurred specifically in the treatment group over the same time periods. Therefore, a valid concern would be that any change in decision patterns, either polarization or moderation, is correlated with the requested amounts. If this is the case, a differential change may occur in the treatment group with regards to other groups, but it is not expected that this effect will be consistent across different control groups.

[14] In Appendix table A2, the same specification is estimated using a sample that excludes "simplified arbitration" proceedings (cases below 25k), to show that the results are not driven specifically by these cases.

Therefore, this alternative explanation can be ruled out if *the same* differential effects are estimated when we use each of the two control groups separately (as presented in Appendix table A2).

Moreover, changes that occur specifically in the treatment group are expected to be portrayed in its composition relative to the comparison group, assuming that there is at least some correlation between observed and unobserved characteristics. To test for differential changes in group composition, the DD specification in equation (1) is estimated using each of the covariates as an outcome separately. The DD coefficient estimates for this set of regressions are reported in Table 1. Out of the 39 characteristics that are tested only two differences are significant at conventional levels and one more is marginally significant. This is roughly what would be expected even if assignment to treatment was random. Nevertheless, the entire set of covariates is controlled for when treatment effects are estimated. Table 1 also presents the mean of each covariate for the sample of cases that are used in the main analysis.

An additional threat to identification occurs when litigants are able to manipulate the rules. In the FINRA setting, claimants actually determine the relief requested and can potentially alter this amount to affect the number of arbitrators. To assess whether claimants manipulate the requested amounts, their distribution was plotted separately for the pre- and post-periods (presented in Figure 4). A Kolmogorov-Smirnov test for equality of distributions was performed to formally compare the two kernel-density plots. The null hypothesis that the samples are drawn from the same distribution cannot be rejected (with a P-value of 0.404), which suggests that claimants did not change their behavior in response to the rule change.

To further confirm that manipulation does not affect the analysis, I redefine the treatment and control groups under the assumption that any claim with a requested relief that precisely equals the threshold amount is a manipulation and therefore it is assigned to the opposite group (namely, if according to the rules, this is a treatment case, it will be redefined as control and vice versa). Results (presented in Table 5) hold after this redefinition, suggesting that the findings are not driven by manipulation.

One more concern is that parties can manipulate the number of arbitrators by changing the filing date. The rule change was announced on February 2009, almost two months before the new rule came into effect, so claimants could either rush to file their claims during this two-month period or delay filing in order to effectively select the number of arbitrators that were assigned to their case. If claimants manipulated the filing date around the time of the policy change, one would expect to see some irregularity in the number of cases filed by month. However, the trends in number of new claims filed during the year of the change appear to be practically identical when the affected group of cases is compared to the control group. Table 2 presents the proportion of cases that were filed during the quarter that preceded the effective date of the new rule out of cases

Table 1
Descriptive Statistics and Group Composition over Time

Variable	Mean	DD
Number of Claimants	1.379	-.13
		(.087)
Claimant Represented	.681	-.03
		(.043)
Number of Respondents	1.734	-.109
		(.138)
Respondent Represented	.981	.035**
		(.014)
Counter Claim	.02	.011
		(.016)
Third Party	.019	-.012
		(.02)
Security type -		
Government Bonds	.024	.005
		(.017)
Corporate Bonds	.067	.004
		(.031)
Funds	.154	.029
		(.045)
Real Estate	.011	-.006
		(.012)
CMOs and MBOs	.006	.009
		(.008)
Certificates of Deposit	.004	-.012
		(.011)
Mutual Funds	.126	.113***
		(.038)
Options	.03	.003
		(.021)
Limited Partnerships	.007	-.009
		(.007)
Annuities	.091	.002
		(.034)
Preferred Stocks	.039	.024
		(.025)
Auction Rate Securities	.019	-.018
		(.012)
Other	.043	-.021
		(.026)
Various Securities	.224	-.08
		(.054)

Variable	Mean	DD
Controversy type -		
Margin Calls	.025	.002
		(.016)
Churning	.049	-.013
		(.025)
Excessive Trading	.137	.006
		(.038)
Failure to Supervise	.401	.053
		(.058)
Negligence	.518	.009
		(.058)
Omission	.271	.047
		(.053)
Breach of Contract	.384	.057
		(.059)
Breach of Fiduciary Duty	.578	.092*
		(.053)
Unsuitability	.35	.006
		(.059)
Misrepresentation	.412	.025
		(.059)
Fraud	.286	.022
		(.056)
Failure to execute	.111	.005
		(.033)
Concentration	.026	.021
		(.018)
Promissory Estoppel	.011	-.014
		(.014)
Elderly Abuse	.008	.012
		(.015)
Respondeat Superior	.14	.056
		(.046)
Unjust Enrichment	.02	-.024
		(.016)
Conversion	.016	.02
		(.019)
Deceptive Practices	.03	.023
		(.024)

Notes: The table reports means for all observed case characteristics and coefficient estimates for the *difference-in-differences* estimation in equation (1) comparing each characteristic pre- and post-treatment for the treatment versus comparison group. The sample is restricted to avoid censoring (as explained in Section 2.2). Standard errors are are clustered at the arbitrator level and reported in parenthesis (*** $p<0.01$, ** $p<0.05$, * $p<0.1$).

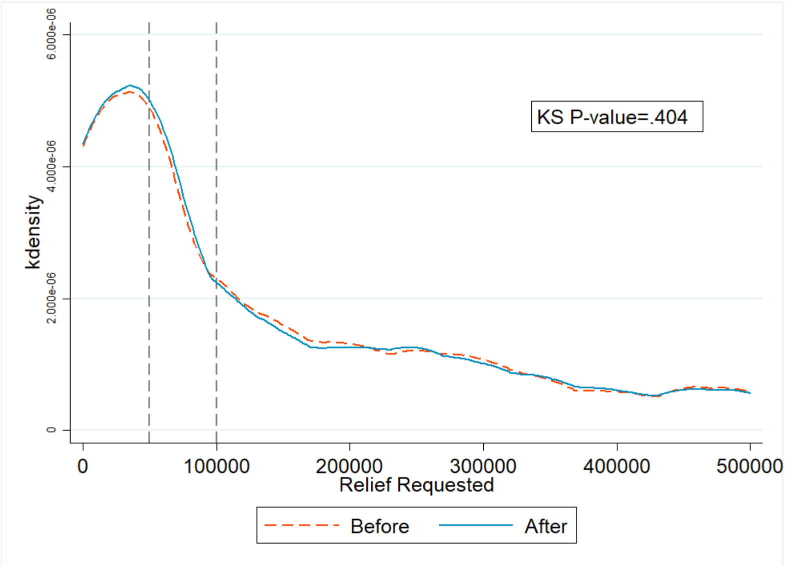


Figure 4
Distribution of relief requested pre- and post-change

Notes: The figure presents kernel density plots of the amount of compensatory damages that were requested by the claimant, before and after March 29, 2009 (pre- and post-the rule change). The p-value for a Kolmogorov-Smirnov test for equality of distribution is presented on the top right-hand corner of the figure (KS P-value). The dashed vertical lines indicate the old and new threshold, \$50k and \$100k respectively. The sample is restricted to avoid censoring (as explained in Section 2.2).

Table 2
Testing for manipulation of filing date

Proportion Filed Pre-change	Treatment (1)	Control (2)	Difference (3)
(a) Jan-Jun, 2009	.472	.496	-.024
			(0.077)
(b) Jan-Dec, 2009	.248	.261	-.013
			(0.049)

Notes: The table reports the share of cases that were filed in 2009 before the regulatory change came into effect, namely until March 29. In row (a) the share is calculated out of the total number of cases that were filed during the first half of 2009 (so that the pre-change period is approximately half of the period). In row (b) the share is calculated for all cases filed during 2009. These shares are reported separately for the treatment and comparison groups. Column (3) presents the difference between the two groups and the standard error. The sample is restricted to avoid censoring (as explained in Section 2.2).

filed during the first half of 2009 and out of cases filed throughout the entire year, by treatment status. In addition, the difference in means between the groups is reported for both measures and is found to be very small and statistically insignificant.

To further establish that there is no manipulation of filing date, Figure 5 shows the number of cases filed each day around the effective date of the new rule, for a period of four months. The data is presented for the whole sample and for each sub-group separately. No significant irregularities were observed during this period.

This finding is actually not surprising since claimants' incentives to manipulate the rules are significantly reduced due to the fact that parties can agree to change the size of the arbitral tribunal and moreover, since prior to the change the consent of the respondent was not always required. However, allowing the litigants to agree on a different number of arbitrators than the regulations set means that compliance with the rule might be partial. Generally, compliance rates are very high in the sample, with more than 95% complying with the rules for the comparison group and 89% for the treatment group.

For the DD estimation strategy, treatment and control were defined according to the expected number of arbitrators (which is set by the rules) in order to avoid bias caused by parties' self selection of the tribunal size. Hence, the estimated parameters should be thought of as intention to treat effects (ITT).

To derive the treatment effect on the compliers, the DD interaction term - $Treated \times Post$ - is used as an instrument for the actual treatment (i.e., sole arbitrator). This is in accordance with the standard practice of using exogenous expected values to instrument for endogenous actual values.

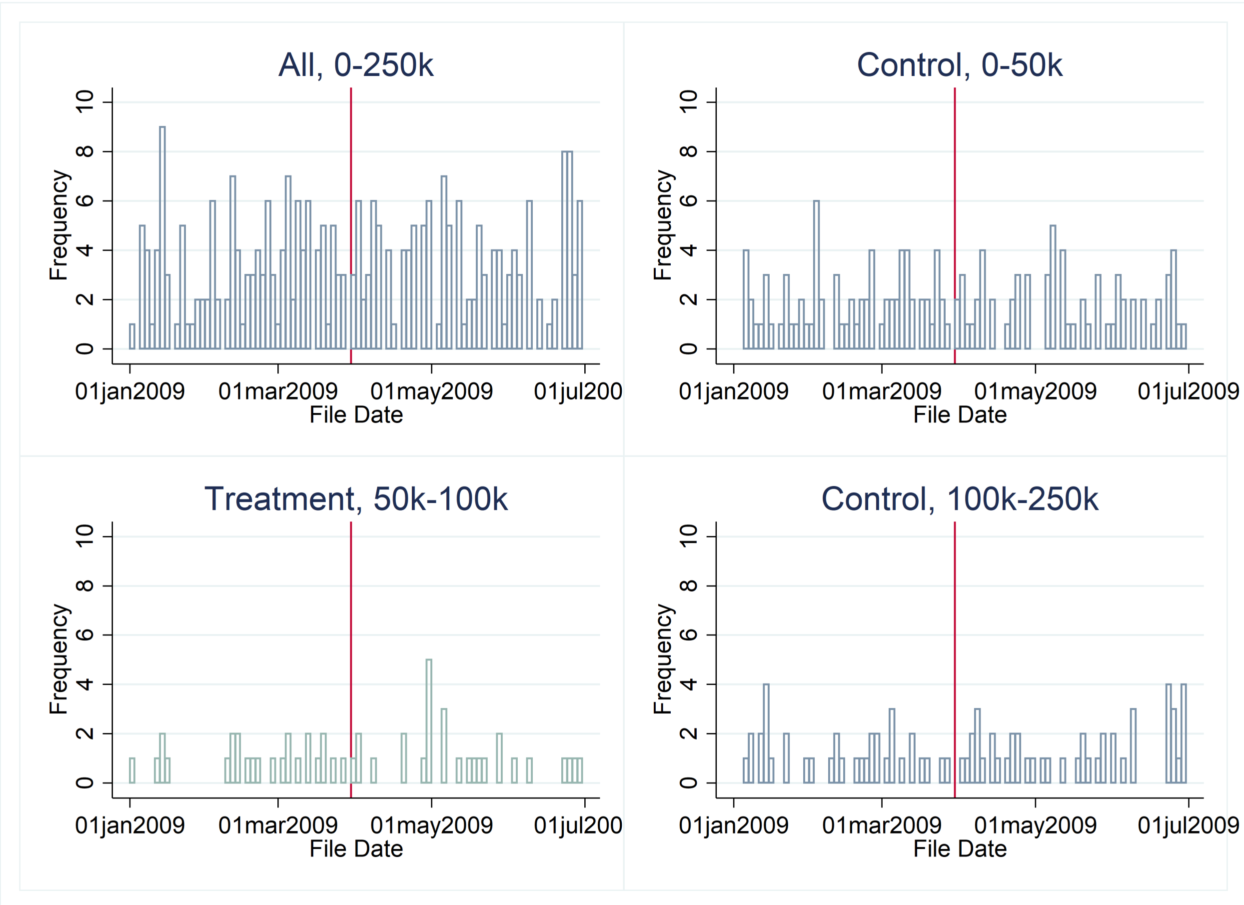


Figure 5
Claim filing frequency pre- and post-change (by range of relief requested)

Notes: The figure presents the number of cases that were filed each day between January 1, 2009 and June 30, 2009, for the entire sample and for each category of cases separately. The vertical line indicates the day that the rule change came into effect (March 30, 2009). The sample is restricted to avoid censoring (as explained in Section 2.2).

Compliers in this setting are cases that are assigned to the same number of arbitrators that FINRA regulations prescribe, and that would have behaved differently if the rules were different. The suggested instrument is highly correlated with the actual number of arbitrators that are assigned because, as mentioned above, non-comliance is rare. Moreover, the instrument is not expected to impact the award rate other than through its effect on the number of arbitrators. Therefore, the instrument is valid and the estimated effect corresponds to the local average treatment effect of having a sole arbitrator decide on the AR (rather than three arbitrators).

4 Results and Discussion

Table 3 reports the estimated coefficients on the interaction term in the DD specification, with and without arbitrator fixed-effects. The sample for the fixed-effect model is smaller since singletons are dropped. Columns (1)-(2) present the estimated change in the proportion (or probability) of extreme awards due to the switch from three arbitrators to a sole arbitrator. The estimates are negative and significant, implying that sole arbitrators are less likely to award either claimants or respondents with the entire claim. The magnitude of the effect remains similar *within* arbitrators.

At the same time, a positive and significant change occurs for moderate or "split the difference" awards, as reported in columns (3)-(4). Note that the decrease in extreme awards is almost fully offset by the increase in moderate awards. This is not necessarily expected since changes could potentially occur for other ranges of AR. Appendix Table A1 presents the same estimates using a full set of indicators for AR value ranges. It is clear that the only significant changes occur either for extreme values or exactly in the middle of the range values. In addition, the average AR is not affected by the change in the number of arbitrators, as shown in column (7) of Appendix Table A1.

Quantitatively, the findings suggest that an individual arbitrator is 20 percentage points less likely to entirely deny or fully accept a claim compared to a panel of three arbitrators. Out of a baseline level of 63%, this constitutes approximately a 32% decrease in extreme awards for sole arbitrators. Appendix table A2 confirms that similar effects are found when the treatment group is compared either to the comparison group with claims below the old threshold (0-50k) or to the group with claims above the new threshold (100-250k). In addition, the findings are robust to excluding claims below 25k which are subject to the simplified arbitration proceedings, although significance levels drop below conventional levels for the arbitrators fixed-effect estimation (probably due to the small sample size). As explained above, the fact that the findings are consistent across these various control groups, rules out any alternative explanation that relies on the association between the requested amount and the tendency of arbitrators to polarize decisions.

In columns (5)-(8), the extreme outcomes are divided according to the prevailing party, either

Table 3
The Effect of Sole Arbitrators on Award Rates

	Extreme		Moderate		Inclination of Extreme Decisions: Respondent (Firm)		Inclination of Extreme Decisions: Claimant (Customer)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TreatedXpost	-0.205***	-0.219**	0.195***	0.182**	-0.107*	-0.130	-0.0980***	-0.0885*
	(0.0585)	(0.0928)	(0.0460)	(0.0711)	(0.0594)	(0.0938)	(0.0312)	(0.0535)
Controls	√	√	√	√	√	√	√	√
Arbitrator FE		√		√		√		√
Observations	1,920	1,255	1,920	1,255	1,920	1,255	1,920	1,255

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports difference-in-differences estimates. The dependent variables are indicators for extreme or moderate awards. In columns (5)-(8) the dependent variable also indicates the prevailing party (claimant or respondent). All specifications include controls for case characteristics (as listed in Table 1)) and filing quarter fixed-effects. The sample is restricted to avoid censoring (as explained in Section 2.2). Standard errors clustered at the arbitrator level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

the claimant or the respondent. For both sides, it is apparent that sole arbitrators are less likely to grant extreme awards, although in some of the specifications significance levels are marginal. According to these estimates, the decrease in the likelihood of extreme awards both in favor of the respondent or in favor of the claimant is approximately 10 percentage points. However, comparing this change to initial levels reveals that while respondents experience a 17% decrease in extreme wins, claimants' extreme wins are practically eliminated (they decrease by almost 100%).

4.1 The Group-Reputation Mechanism

The results in Table 3 are based on the sample of customer cases in which the claimant is always the investor (customer) and the respondent is a firm. In such cases, litigants may suspect arbitrators for being biased towards either side in a systematic way, based on each arbitrator's past decisions. Since arbitrators are penalized for extreme decisions by gaining a "bad" reputation as biased, some arbitrators may try to refrain from extreme awards.

It was already established that differential selection of arbitrators into panels cannot drive the effects that are found within arbitrators. The group-reputation mechanism, on the other hand, predicts that *the same* arbitrator will be more cautious and grant more moderate awards if she decides alone rather than with a panel. This is because panel decisions send a much noisier signal of individual arbitrators' potential biases than sole arbitrators' decisions. In addition to being in line with the findings that panels tend to be more extreme, this mechanism offers an adequate explanation to the asymmetrically larger impact on extreme customer wins. In the FINRA setting it is likely that the reputational penalty for extreme decisions in favor of customers is substantially higher than for extreme decisions in favor of firms who are repeat players in arbitration and as such enjoy an informational advantage over customers (Egan et al., 2018).

Still, an alternative explanation to the findings could be that certain cognitive biases are either exacerbated or attenuated in groups, and this impacts their decisions. However, if this is the mechanism that drives the polarization of panel awards in customer cases, we should expect the same polarization to occur in other cases, namely industry cases. In contrast, the group-reputation mechanism would only be expected to operate in the presence of reputation concerns. Hence, it should not come into play in industry cases, in which both sides to the dispute are usually firms and arbitrators' reputation cannot be negatively affected by extreme decisions. The assertion that reputation concerns are less important in industry cases is supported by the fact that, regardless of the number of arbitrators, the proportion of extreme awards in industry cases is 13 percentage points larger than in customer cases. In addition, FINRA arbitrators' impartiality is often discussed when it comes to disputes between firms and customers, but not with regards to other types of disputes,[15] and in a recent study Egan et al. (2018) focus specifically on customer cases to show that parties, especially firms, consider past decisions of arbitrators as a significant signal for potential bias.

Table 4 presents the estimated differences between panels and sole arbitrators for industry cases. Industry cases were subject to the same change in regulations that is used to identify the difference between individuals and groups in the sample of customer cases. Hence, the exact same equation is estimated (equation (1)), and the same sample restrictions are used to avoid data censoring. The only difference is that industry cases display immense variation in controversy issues and hence the causes of claim and subjects of dispute could not be categorized and controlled for as with customer cases.

The results for industry cases do not resemble the findings for customer cases. No distinct pattern of moderation or polarization can be identified and none of the estimates are significant. These findings strongly support the proposed group-reputation mechanism as the most probable

[15]For example, in an article that was published in their website, security lawyers inform their potential clients about the "wide-spread perception that the arbitration system is stacked against customers" and the "concern that many arbitrators are reluctant to mete out large awards against big brokerage firms". http://www.sbpllplaw.com/an-outline-of-the-finra-arbitration-process-for-customer-broker-disputes/. On the contrary, an article by Holtzman, Harlow & Berkley in the New York Law Journal discusses brokerage firms' efforts to *avoid* the "customer-friendly" FINRA mandatory arbitration (published March 30, 2015).

Table 4
The Effect of Sole Arbitrator on Award Rates in Industry Cases

	Extreme		Moderate	
	(1)	(2)	(3)	(4)
TreatedXpost	0.0258 (0.0549)	-0.115 (0.105)	0.00169 (0.0299)	-0.00594 (0.0576)
Controls	√	√	√	√
Arbitrator FE		√		√
Observations	1,239	544	1,239	544

Notes: The table reports difference-in-differences estimates for a sample of FINRA industry cases. The dependent variables are indicators for extreme or moderate awards. All specifications include filing quarter fixed-effects. The sample is restricted to avoid censoring, using the same rules as for the main sample (as explained in Section 2.2). Standard errors clustered at the arbitrator level in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

channel that explains the systematic differences between group and individual decisions. Behavioral explanations to the polarization of groups cannot account for the contradicting conclusions regarding customer and industry cases, namely that panels polarize decisions in customer cases but have no significant impact on decisions in industry cases.

4.2 Robustness of Results

To establish that the main findings for customer cases are not driven by events other than the rule change or by the way the data and the empirical analysis were structured, in this section I present the results of various robustness and placebo tests. Te first issue is that the results presented thus far are only based on a sample of approximately 90% of all cases. This is because I only use awards that are given within a specific period since filing due to the retrospective nature of the data. To see how this might affect results, I conduct the following experiment. I use all the awards in the full sample for the pre-period, assuming that the pre-change sample is complete. This assumption is reasonable since starting from the latest filing date, awards are collected for almost four additional years, and the number of cases that are expected to resolve after more than four years is negligible.

For the post-period, I add artificial observations that simulate the "worst case" scenario for the prediction of interest. More specifically, the additional control group observations are assumed to follow the pattern predicted for the *treatment* group in the post-period, namely all awards are moderate. The opposite is done for the treatment group in that all of the added observations are coded as extreme awards. Row (b) of Table 5 presents the estimated effects on extreme and moderate awards using this extended sample. When compared to the main results (presented again for convenience in row (a)), these estimates should be thought of as the *lower bound* of the impact. Clearly, the point estimates exhibit the same trends although with smaller magnitudes of change (as expected). Significance levels are maintained. Keeping in mind that this scenario is extreme and highly unlikely, the experiment indicates that results will hold under all possible scenarios for the missing data.

Rows (c)-(f) in Table 5 display an array of additional robustness tests, the results of which reinforce the findings that were presented in Table 3. First, in rows (c) and (d) I test whether the results are sensitive to changes in the case duration limit (aimed at avoiding censoring effects). Row (c) presents the results for a sample without any restriction on case duration,[16] while row (d) shows the estimates when a more stringent restriction is applied, i.e., the maximal number of days is the 75th percentile of the case duration distribution. For the most part, the choice of maximal duration has no effect on results. The only exception is the magnitude of the effect on moderate awards in the fixed-effects specification where the effect is slightly smaller (but still substantial) and only marginally significant. However, this is expected with such a small sample size (less than 1,000 observations).

Next, in row (e), I restrict the sample to exclude all cases filed during 2009, the year of the regulatory change. The estimated effects of the number of arbitrators for this sample are actually larger and significant, although the sample size is much smaller. The aim of this test is twofold. First and foremost, because most claims that relate to the events of the financial crisis were filed during 2009, omitting this year from the analysis confirms that the main findings are unrelated to effects of the crisis on FINRA arbitration (such as new types of claims or negative attitudes towards financial markets). Second, if there was a concern that some manipulation of filing dates occurred around the time of the rule change, this analysis excludes the manipulated cases and shows that results hold.

In row (f) the treatment and control groups are redefined to account for possible manipulation of claims around the 50k and 100k cutoffs. This test supplements the findings in Figure 4 that rule out manipulation by showing that the distribution of requested relief was not affected by the rule change. In the main analysis, cases with a relief requested exactly equal to 50k are included in the control group both before and after the change. If we assume that pre-change claimants tended to choose this amount to assure that their case is heard by a sole arbitrator (even though they actually intended to request a slightly higher amount), then, in the pre-period, some of the control group cases should in fact be in the treatment group (over 50k). Similarly, if the claimed amount is precisely 100k, the case is defined as treated, but in the post-period, it could be that the amount

[16]The last filing year 2011 is omitted to avoid severe censoring.

Table 5
Robustness and Placebo Tests

	Extreme		Moderate	
	(1)	(2)	(3)	(4)
(a) Main	-0.205***	-0.219**	0.195***	0.182**
	(0.0585)	(0.0928)	(0.0460)	(0.0711)
Observations	1,920	1,255	1,920	1,255
(b) Lower Bound	-0.179***		0.178***	
	(0.0516)		(0.0423)	
Observations	2,176		2,176	
(c) Unlimited Duration	-0.177***	-0.228***	0.184***	0.186***
	(0.0548)	(0.0850)	(0.0437)	(0.0640)
Observations	1,996	1,325	1,996	1,325
(d) 75th pct Duration	-0.217***	-0.203**	0.183***	0.145*
	(0.0617)	(0.100)	(0.0471)	(0.0750)
Observations	1,586	941	1,586	941
(e) Excluding 2009	-0.275***	-0.244**	0.243***	0.255**
	(0.0686)	(0.115)	(0.0611)	(0.0984)
Observations	1,394	758	1,394	758
(f) Moving Thresholds	-0.210***	-0.197**	0.170***	0.124
	(0.0604)	(0.0951)	(0.0501)	(0.0768)
Observations	1,920	1,255	1,920	1,255
(g) Re-coded settlements	-0.205***	-0.219**	0.102***	0.0872
	(0.0585)	(0.0928)	(0.0351)	(0.0534)
Observations	1,920	1,255	1,920	1,255
(h) Pre-Period Placebo	-0.0308	0.153	0.000916	-0.122
	(0.0788)	(0.158)	(0.0553)	(0.0959)
Observations	1,017	454	1,017	454
Controls	√	√	√	√
Arbitrator FE		√		√

Notes: The table reports difference-in-differences estimates of the impact of the number of arbitrators on extreme and moderate ARs, using various samples and restrictions to test the robustness of the main results (presented for convenience in row (a)). All specifications include controls for case characteristics (as listed in Table 1)) and filing quarter fixed-effects, except for the sensitivity analysis in row (b), where control variables were not generated for the simulated data. The sample is restricted to avoid censoring (as explained in Section 2.2), except in rows (c) and (d). In row (c) duration is not limited but cases that were filed during 2011 are omitted. In row (d) the sample is restricted to cases where the duration of proceedings is lower or equal to the 75th percentile of case duration in each group of cases. Standard errors clustered at the arbitrator level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

was manipulated in order to reduce the number of arbitrators and the case should have in fact belonged to the control group. If this is true, the composition of the groups was affected by the rule change and hence estimates may be biased. In accordance with this possibility, the treatment and control groups are redefined as follows: The 50k cases are moved from the control group into the treatment group only in the pre-period and the 100k cases are moved from the treatment group to the control group only in the post-period. Note that the 50k threshold is meaningless in the post-period and the 100k threshold is meaningless in the pre-period. Re-estimating the main specifications using this new definition of control and treatment does not change the main findings except that the estimated within-arbitrator impact on moderate awards drops slightly below conventional significance levels (with a p-value of 0.105).

Last, in row (g), to deal with the missing data problem caused by the confidentiality of most settlement agreements, instead of simply classifying all settlements as moderate awards, I use the predicted AR for cases that were settled (based on a regression that includes all observable case characteristics) and define only settlements with predicted ARs between 0.4 and 0.6 as moderate. Using predicted values for the outcome variable is generally inadvisable, and therefore the results are only presented as a robustness test. Nevertheless, in this particular instance, assuming that settlement amounts equal predicted awards is supported by the well-known theoretical result that settlements occur when awards are predictable and parties prefer to settle for the predicted award in order to avoid the costs of trial (Priest and Klein, 1984).

When this alternative definition is applied, results regarding extreme awards remain unchanged while the impact on moderate awards is smaller, but still positive and significant. This analysis also implies that about half of the increase in moderate awards for sole arbitrators is due to an increased propensity to settle. This cannot be explained by costs, since settlement rates are expected to increase with the costs of proceedings (Bebchuk, 1984), and hence with the number of arbitrators. However, it can be the result of lower variation that is found in sole arbitrators' awards, namely the fact that they are easier to predict. Note that this result contradicts the findings of Marselli et al. (2015) who report that settlements are associated with the presence of three arbitrators rather than one. However, in their sample panel decisions appear to be more predictable.

To further establish that the results are not coincidental or simply follow a pre-existing trend, a placebo test is performed, the results of which are presented in the last row of Table 5. This placebo test uses the same treatment and control groups but only in the pre-period. This period is divided into two equal sub-periods, setting August 15, 2007 as the placebo-treatment date. None of the effects are significant, point estimates are inconsistent across specifications, and they do not resemble the main result, namely that sole arbitrators' decisions are less extreme. Although the results are reported only for one placebo-treatment date, this exercise was repeated with various random dates, and the findings remain similar. These results are helpful in refuting concerns regarding any difference in pre-trends between treatment and control groups.

4.3 Instrumental Variable Estimation

All the preceding results are based on the expected number of arbitrators which are determined according to FINRA regulations that were applicable on the filing day. Although parties comply with the rules 90% of the time, the lack of full compliance implies that the results displayed so far should be thought of as ITT. Namely, the estimated impacts are a consequence of the rule change and not of the actual forum hearing the case. To assess the effect of the change in the number of arbitrators on the outcome among the compliers, I next present the results of an estimation that uses the expected values of the key variable to instrument the actual values. More specifically, the interaction between the rule change and the affected group is used as an instrumental variable for sole arbitrators.

The proposed instrument satisfies the necessary conditions. First, as established above, the regulatory change can be viewed as exogenous to the outcomes of interest. Second, it is reasonable to assume that the exclusion restriction holds since there is no reason to believe that the rule change for the treatment group affected arbitration outcomes by any other channel, except for the change in the number of arbitrators. Third, columns (1)-(2) of Table 6 present the results for the first stage (with and without arbitrator FEs) and confirm that the instrument is highly correlated with the appointment of a sole arbitrator. In addition, the F-statistic for the first stage is larger than 190. Thus, the 2SLS estimators reported in columns (3)-(6) are consistent and, as expected, they are larger in absolute values compared to the DD estimates (reported in Table 3 above) and are highly significant. This reinforces the conclusion that panels and sole arbitrators behave and decide differently.

5 Conclusion

This study presents empirical evidence on the difference between decisions of sole arbitrators and panels of three arbitrators. The results indicate a tendency of panels to make more extreme decisions compared to sole arbitrators. In addition, panels seem to be less inclined to grant 'split-the-difference' awards or to promote settlements. I show that these findings are robust and most likely driven by the fact that reputation concerns affect sole arbitrators, driving them to avoid extreme outcomes that can potentially establish their reputation as biased towards either side. Alternatively stated, when arbitrators decide as a group they exhibit less concern for their reputation

Table 6
The Effect of Sole Arbitrator on Award Rates - IV Approach

	1st Stage		2SLS Extreme		2SLS Moderate	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated×Post	0.716***	0.706***				
	(0.0347)	(0.0523)				
Sole public arbitrator			-0.283***	-0.304**	0.269***	0.253**
			(0.0804)	(0.131)	(0.0639)	(0.0998)
Controls	√	√	√	√	√	√
Arbitrator FE		√		√		√
Observations	1,920	1,255	1,920	1,255	1,920	1,255

Notes: The table reports IV estimates, for the main outcome variables, using the interaction term from the DD specification as an instrument for the endogenous explanatory variable - sole arbitrator. Columns (1)-(2) Report the results of the first stage estimation. Columns (3)-(6) present the second stage estimates. All specifications include controls for case characteristics (as listed in Table 1)) and filing quarter fixed-effects. The sample is restricted to avoid censoring (as explained in Section 2.2). Standard errors clustered at the arbitrator level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

and hence have the courage to make extreme and perhaps even controversial decisions.

However, I cannot evaluate the welfare consequences of the systematic differences between group and individual decisions. While the actual presence of reputation concerns of the type that was described here implies that some arbitrators must be biased, it is impossible to distinguish them from impartial arbitrators or to estimate their proportion. Holding the level of reputation concerns constant, if only a handful of arbitrators are actually biased, it is plausible to assume that on-average sole arbitrators make *less* accurate decisions. However, as the proportion of biased arbitrators decreases, litigants will express less interest in estimating arbitrators' bias based on past decisions and therefore arbitrators should be much less worried about making extreme decisions. Following this intuition, the substantial impact that this study reports implicitly suggests that a sizable proportion of the arbitrators are biased, and hence it is more likely that panel decisions are on average less accurate and more biased.

While the conclusions of this study can be extended to other contexts and settings where decisions are made either by groups or by individuals, this should be done with caution. This is because, theoretically, different mechanisms are expected to operate conditional on the type of decisions, on group characteristics (such as size and diversity), and on decision procedures (e.g. extent of deliberation, voting rules). Given the prevalence of group decisions in various environments and the vast interest in this subject in multiple disciplines, such as social psychology, management, law and political science, finding more opportunities to empirically identify the differences between group and individual decisions is essential.

References

Adams, R. and D. Ferreira (2010). Moderation in groups: Evidence from betting on ice break-ups in alaska. *The Review of Economic Studies 77*(3), 882–913.

Arlen, J. and S. Tontrup (2015). Does the endowment effect justify legal intervention? the debiasing effect of institutions. *The Journal of Legal Studies 44*(1), 143–182.

Bazerman, M. H. and H. S. Farber (1985). Arbitrator decision making: when are final offers important? *Industrial & Labor Relations Review 39*(1), 76–89.

Bebchuk, L. A. (1984). Litigation and settlement under imperfect information. *The RAND Journal of Economics*, 404–415.

Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets. *The Review of Economic Studies 80*, 429–462.

Charness, G. and M. Sutter (2012). Groups make better self-interested decisions. *The Journal of Economic Perspectives 26*(3), 157–176.

Cooper, D. J. and J. H. Kagel (2005). Are two heads better than one? team versus individual play in signaling games. *The American Economic Review 95*(3), 477–509.

Da, Z. and X. Huang (2018). Harnessing the wisdom of crowds. *forthcoming in Management Science*.

Egan, M. L., G. Matvos, and A. Seru (2018). Arbitration with uninformed consumers. *Working paper*.

Eliaz, K., D. Ray, and R. Razin (2006). Choice shifts in groups: A decision-theoretic basis. *The American Economic Review 96*(4), 1321–1332.

Ely, J. C. and J. Välimäki (2003). Bad reputation. *The Quarterly Journal of Economics*, 785–814.

Epstein, L., W. M. Landes, and R. A. Posner (2011). Why (and when) judges dissent: A theoretical and empirical analysis. *Journal of Legal Analysis 3*, 101–137.

Farber, H. S. (1981). Splitting-the-difference in interest arbitration. *Industrial & Labor Relations Review 35*(1), 70–77.

Farber, H. S. and M. H. Bazerman (1986). The general basis of arbitrator behavior: An empirical analysis of conventional and final-offer arbitration. *Econometrica 54*(4), 819–844.

Fraser, C., C. Gouge, and M. Billig (1971). Risky shifts, cautious shifts, and group polarization. *European Journal of Social Psychology 1*(1), 7–30.

Glaeser, E. L. and C. R. Sunstein (2009). Extremism and social learning. *Journal of Legal Analysis 1*(1), 263–324.

Goeree, J. K. and L. Yariv (2011). An experimental study of collective deliberation. *Econometrica 79*(3), 893–921.

Janis, I. L. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Houghton Mifflin Boston.

Kerr, N. L., R. J. MacCoun, and G. P. Kramer (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review 103*(4), 687–719.

Klement, A. and Z. Neeman (2013). Does information about arbitrators' win/loss ratios improve their accuracy? *The Journal of Legal Studies 42*(2), 369–397.

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

Marselli, R., B. C. McCannon, and M. Vannini (2015). Bargaining in the shadow of arbitration. *Journal of Economic Behavior & Organization 117*, 356–368.

Meade, E. E. and D. Stasavage (2008). Publicity of debate and the incentive to dissent: Evidence from the us federal reserve. *The Economic Journal 118*(528), 695–717.

Morris, S. (2001). Political correctness. *Journal of Political Economy 109*(2), 231–265.

Moscovici, S. and M. Zavalloni (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology 12*(2), 125–135.

Ottaviani, M. and P. Sørensen (2001). Information aggregation in debate: who should speak first? *Journal of Public Economics 81*(3), 393–421.

Posner, R. A. (2005). Judicial behavior and performance: An economic approach. *Fla. St. UL Rev. 32*(4), 1259–1280.

Posner, R. A. (2010). *How Judges Think*. Harvard University Press.

Prat, A. (2005). The wrong kind of transparency. *American Economic Review 95*(3), 862–877.

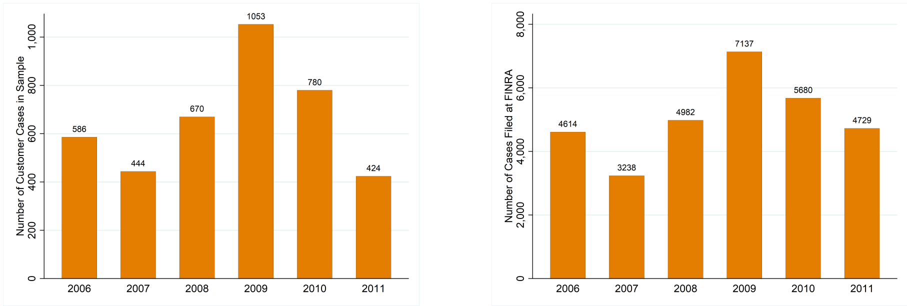
Priest, G. L. and B. Klein (1984). The selection of disputes for litigation. *The Journal of Legal Studies 13*(1), 1–55.

Stoner, J. A. (1968). Risky and cautious shifts in group decisions: The influence of widely held values. *Journal of Experimental Social Psychology 4*(4), 442–459.

Stoner, J. A. F. (1961). *A comparison of individual and group decisions involving risk*. Ph. D. thesis, Massachusetts Institute of Technology.

Zamir, E. and D. Teichman (2018). *Behavioral Law and Economics*. Oxford University Press.

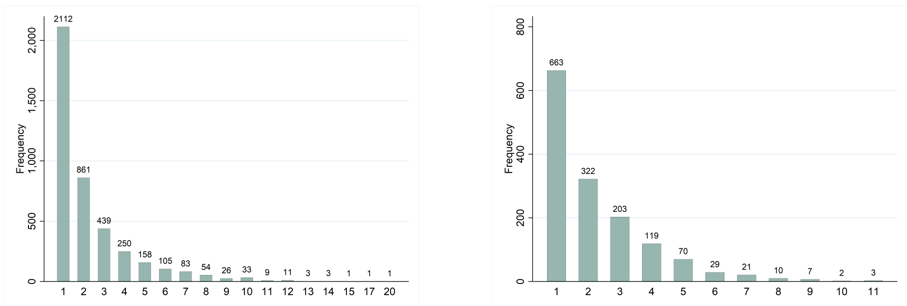
6 Appendix



(a) Awards in Customer Cases (b) All Cases Filed

Figure A1
Number of Cases by Filing Year

Notes: Figure (a) on the left-hand side presents the number of cases filed in FINRA arbitration for each year, between 2006 and 2011. The source for this data is the statistics posted by FINRA on their website https://www.finra.org/. Figure (b) on the right-hand side, shows the number of customer cases in the full sample collected for this study, by filing year. Due to the retrospective nature of the data, only cases that were closed by January 2013 are counted, and this probably explains the relatively small number of cases in 2011.



(a) Any Position, All Customer Cases (b) Chair or Sole, in Sample

Figure A2
Number of Cases per Arbitrator

Notes: The two figures present the frequency of number of times that the same arbitrator appears in the data. Figure (a) on the left-hand side presents the number of times that the same arbitrator appears in the sample in any position — sole arbitrator, panel-chair or panel-member. Figure (b) on the right-hand side presents the number of times that the same arbitrator appears in the sample as sole or as chair.

Table A1
The Effect of Number of Arbitrators on Award Rate Distribution

	(1) $0 \leq AR < .2$	(2) $.2 \leq AR. < 4$	(3) $.4 \leq AR < .6$	(4) $.6 \leq AR < .8$	(5) $.8 \leq AR < 1$	(6) $AR \geq 1$	(7) Avg. AR	*FE*	*Obs.*
(a) TreatedXpost	-0.123**	0.0217	0.195***	0.00503	-0.000806	-0.0980***	-0.0139		1920
	(0.0583)	(0.0323)	(0.0460)	(0.0267)	(0.0235)	(0.0312)	(0.0537)		
(b) TreatedXpost	-0.155*	0.0156	0.182**	-0.00274	0.0492	-0.0885*	-0.0109	$\surd$	1255
	(0.0869)	(0.0406)	(0.0711)	(0.0285)	(0.0351)	(0.0535)	(0.129)		

Notes: The table reports difference-in-differences estimates. The dependent variables are indicators for the AR being in the range specified in the column title. All specifications include controls for case characteristics (as listed in Table 1)) and filing quarter fixed-effects. The sample is restricted to avoid censoring (as explained in Section 2.2). Standard errors are clustered at the arbitrator level and reported in parenthesis (*** p<0.01, ** p<0.05, * p<0.1).

Table A2
The Effect of Sole Arbitrators on Award Rates — Alternative Control Groups

	Control Group					
	0-50K		100-250K		25-50K,100-250K	
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable:						
Extreme Award	-0.230***	-0.174*	-0.164**	-0.286*	-0.174***	-0.173
	(0.0614)	(0.102)	(0.0708)	(0.170)	(0.0661)	(0.139)
Moderate Award	0.237***	0.244***	0.132**	0.286**	0.151***	0.241**
	(0.0469)	(0.0761)	(0.0562)	(0.134)	(0.0524)	
Controls	√	√	√	√	√	√
Arbitrator FE		√		√		√
Observations	1,432	831	849	353	1,040	497

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports difference-in-differences estimates. The dependent variables are indicators for extreme or moderate awards. All specifications include controls for case characteristics (as listed in Table 1)) and filing quarter fixed-effects. The sample is restricted to avoid censoring (as explained in Section 2.2). Standard errors clustered at the arbitrator level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1