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

Naomi Gershoni*

Apr 20, 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 are derived from laboratory experiments, and hence do not reflect professional incentives and career concerns, which may play a crucial role. This paper uses an original data set of arbitration awards from the Financial Industry Regulatory Authority and a unique regulatory change that exogenously decreased the number of presiding arbitrators from three to one for a specific class of cases. A difference-in-differences strategy is implemented to identify the impact of the number of decision-makers on the distribution of outcomes. 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 arbitrator, implying that the same arbitrators are inclined towards more extreme, “all or nothing” decisions, when in groups. This rules out the possibility that the effect is driven by differences in the selection of arbitrators into panels. Rather, evidence support a novel explanation to the polarization of groups — individuals’ concerns about adverse effects of extreme decisions on their reputation are mitigated within groups, where individual opinions are at least partially obscure.

*Ben Gurion University of the Negev
1 Introduction

The importance of understanding the systematic differences between group and individual decisions has been well recognized in the economic literature. Decisions are made by groups in a wide array of issues and settings. Some examples are business decisions by boards, policy recommendations by committees, household choices, and judicial panels or jury decisions. 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 better and that more intricate psychological and economic mechanisms may drive differences between group and individual decisions. Moreover, such differences often depend on context and institutions, as well as on the criteria by which the decision is evaluated. In light of the numerous competing predictions, the scarcity of empirical evidence based on “real world” observational data is striking, though not surprising. 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. Empirical evidence that come from laboratory experiments do not 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, where cases are heard and decided-on by either one arbitrator or three. It contributes to a long standing discussion on the polarization of groups (dating back to Stoner, 1961), by focusing on whether group decisions tend to be more or less extreme than individual decisions. While evidence from laboratories consistently find groups to be more extreme than individuals, the only study I am aware of that systematically compares individual and group decisions in the field finds a moderating effect of groups on outcomes (Adams and Ferreira, 2010). This debate is particularly interesting in the context of arbitrators’ behavior, where it was suggested that avoiding extreme decisions may benefit arbitrators’ reputation as fair and unbiased (Posner, 2004; Klement and Neeman, 2013) and that arbitrators may (or may not) simply choose to “split the difference” between parties, regardless of evidence and facts (Farber, 1981; Farber and Bazerman, 1986; Bazerman and Farber, 1985).

Comparing decisions of sole arbitrators to those of panels to establish causal effects requires a setting where the assignment of specific cases to one option or the other is random, or as good as random. In reality, the number of arbitrators is either determined by an initial agreement between the parties to a dispute or according to a specific rule that assigns different types of cases to different tribunals. Therefore, the results of a “naive” comparison between the decisions taken

---

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

2See Glaeser and Sunstein (2009) for a short review of this literature.
by different forums may actually reflect the underlying difference in the characteristics of the dispute and not in the behavior of arbitrators. To overcome this challenge, I use a unique data set of the Financial Industry Regulatory Authority (FINRA) arbitration awards and a regulatory change that exogenously decreased the number of arbitrators assigned to a specific group of cases.

The study focuses on claims of investors (customers) against brokerage firms (“Customer Cases”) that were filed in FINRA arbitration between 2006 and 2011. 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 the group of cases between the two thresholds. These cases 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, this change was induced by an unusual load of cases during the year 2009, probably following the financial crisis that peaked in the preceding year. While allowing me to consider the rule change as an exogenous shock to tribunal size, this chain of events 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 awards may simply exhibit a time trend. To address these concerns and confounding factors, I use a difference-in-differences framework. Cases that were not affected by the change, namely cases with a relief requested below the old threshold or above the new one, serve as a comparison group. I present elaborate balance tests and a variety of robustness tests to validate this identification strategy. In addition, an Instrumental Variable approach is used to address partial compliance with the rule (although compliance levels are around 90%).

The main outcome that is considered is the award rate, i.e. the ratio of awarded damages to the requested relief. The results show that the distribution of awards significantly changes for the treatment group in a way that is not exhibited by the control group of cases. More specifically, when compared to sole arbitrators, panels of three are more likely to deliver an extreme award, either entirely dismissing claims or awarding claimants with very high amounts of damages relative to their claim. Sole arbitrators tend to render more moderate awards, which grant each side with approximately half of their claim. These results suggest that groups decisions systematically differ from individual decisions, although no inference can be made regarding the quality of the decisions.

Multiple robustness and placebo tests are conducted to confirm these findings. Importantly, the analysis is repeated using a sample that excludes cases from 2009 in order to avoid the impact of the financial crisis. In addition, an Instrumental Variable approach is implemented to deal with partial compliance to the rules. In both cases, estimates point to the same direction of change and
are actually larger in magnitude.

To further assess the implications of these findings, I consider and evaluate three potential underlying mechanisms. First, the selection of arbitrators into panels may be different and hence the difference in outcomes could be attributed to the difference in characteristics of arbitrators that serve as sole arbitrators and those who serve in panels. I rule out this option by using an arbitrator fixed-effects model and showing that an effect of a similar magnitude is found within arbitrator.

Second, cognitive biases, which were proposed in order to explain group polarization in laboratory experiments, could drive the portrayed differences. If this is the case, the normative implication is that group decisions are less accurate, or that the distribution of individual decisions is more similar to the distribution of correct outcomes. This is especially true if some arbitrators are biased and if there is a tendency to group in panels arbitrators with similar views.

Alternatively, the same findings could be the result of a perfectly rational choice by arbitrators with reputation concerns. In FINRA arbitration, litigants’ preferences over arbitrators affects their selection and in customer cases, parties can try to learn about arbitrators’ bias based on past decisions in similar cases (namely, other customer cases). Importantly, while previous decisions of arbitrators and their identities are easily accessible to all, the truth (namely the correct decision) remains private information. In this framework, extreme decisions could be interpreted as portraying favoritism and hence, arbitrators may prefer to avoid such decision even if the facts of the case support them. Under the plausible assumption that group decisions at least partially obscure individual opinions and therefore have less of an impact on arbitrators’ reputation, groups will be less reluctant than individual arbitrators to grant extreme awards.

In order to distinguish the last two channels, I repeat the main analysis on “Industry Cases”, where customers are not a side to the dispute, and hence reputation is less of a concern. In this “reputation free” setting, no significant differences are found between groups and individuals, thus suggesting that the most likely explanation to the group extremism identified in customer cases, is the mitigation of reputation concerns within groups. While this implies that groups potentially exacerbate bias when arbitrators are not neutral, groups also facilitate truthfulness of impartial arbitrators when the facts warrant an extreme award.

In addition to the general contribution of this paper to the analysis of group decision-making, the findings may have policy implications in the specific context of arbitration proceedings. As arbitration is becoming an increasingly widespread alternative to courts, discussing the institutional aspects of arbitration becomes essential. Contrary to courts, the number of arbitrators assigned to a case is usually determined by the parties to the dispute. The common practice is to increase the number of arbitrators as the stakes of the case increase. while choosing more arbitrators imposes higher costs of proceedings, the benefit of this choice is not clear, as none of the (very few)
empirical studies that tackle the difference between group and individual decisions address arbitration decisions. On the other hand, the large body of literature on arbitrators’ decision-making have always focused either on decisions of a single arbitrator or on decisions of panels, but the two have never been compared. The only exception to this is Marselli et al. (2015) that consider the association between the number of arbitrators and the probability of a dispute resolving in settlement.

The paper proceeds as follows— The next section briefly surveys related literature and theoretical predictions. In section 3 the FINRA Arbitration setting is reviewed in detail. Next, I describe the unique data set and the empirical strategy. Results and robustness tests are presented in section 4. Section 5 tests the plausibility of the alternative mechanisms and presents evidence in support of the “reputation channel”. Section 6 concludes.

2 Related Literature and Conceptual Framework

Previous studies of group decision-making suggest various mechanisms that distinguish between group and individual decisions but no consistent predictions or conclusions emerge. While some studies emphasize the moderating effect of groups on potentially extreme individual opinions, others present the possibility that groups evoke polarization of individual views and motivate manipulation of outcomes.

The well-known Condorcet Jury Theorem, predicts that majorities are more likely than any single individual to select the best out of two alternatives, based on the assumption that information aggregation is efficient. Experimental studies support this prediction by repeatedly finding group decisions to be more consistent with rationality (Kocher and Sutter, 2005; Baillon et al., 2016; Bornstein and Yaniv, 1998; Rockenbach et al., 2007; Blinder and Morgan, 2005).

In addition, it was claimed that groups are more predictable and less extreme due to a ‘compromise effect’— individual members’ opinions averaging out in order to reach consensus (Adams and Ferreira, 2010). Even when consensus is not required, studies document a tendency towards unanimity or ‘dissent aversion’, especially when members of the group interact repeatedly, e.g.

The vast majority focus on investment decisions and offer mixed evidence for the association between groups and outcomes (Barber et al., 2003; Barber and Odean, 2000; Bone et al., 1999; Prather and Middleton, 2002). Adams and Ferreira (2010) compare the distribution of bets placed by individuals and groups on ice breakup dates and find groups to be more moderate and conform more to historic data. However, the setting of a specific wagering game is significantly different from other economic and judicial decisions.

The seminal papers in this string of literature date back to the 80s. For a specific discussion of the determinants of decisions in FINRA arbitration see Choi et al. (2010).

A long line of literature that followed altered and relaxed the underlying assumptions and a variety of results were derived (see Piketty (1999) for a survey of papers in this line of literature).
panels of judges (Posner, 2010). Moreover, Adams and Ferreira (2010) predict that the ‘membership effect’ – groups choosing to exclude individuals with extreme points of view or such individuals voluntarily leaving groups – will amplify the moderating effect of groups. Since usually arbitrators are chosen by the litigants, it is plausible that some type of ‘selection effect’ will be present, namely that specific types of arbitrators are chosen to serve in panels while others are more likely to serve as sole arbitrators, though it is difficult to predict if this is a moderating or a polarizing effect.

However, others cast doubt on the wisdom and moderation of groups. Social psychology points to several cognitive biases which may lead to the opposite outcome. The phenomena known as group polarization, was repeatedly documented in lab experiments and was supported by a variety of theoretical models (Glaeser and Sunstein, 2009). Two types of potential choice shifts of groups, either “risky” or “cautious”, were demonstrated under divergent circumstances and shown to depend on group members’ individual predispositions and the extent of deliberation (see e.g. Stoner, 1968, 1961; Moscovici and Zavalloni, 1969; Kerr et al., 1996). An even more alarming view of group behavior was first described by Janis (1982), who coined the term “group-think” to describe the tendency of groups to apply self-deception in an attempt to conform with each other and avoid conflict.

Ottaviani and Sørensen (2001) stress on a different aspect of group decisions– reputation concerns lead to flaws in information aggregation. They predict that when committee members present their opinions sequentially, they will end-up adopting relatively extreme views, a phenomenon known as ‘herding’. Levy (2007) analyzes decisions made by committees whose members are motivated by similar concerns for reputation and focuses on the interaction of these concerns with the transparency of the decision process. In this setting, transparency promotes decisions that go against members’ preexisting bias, whereas in others it may promote more conformity (Meade and Stasavage, 2008).

When the decision process is not transparent, as with panels of arbitrators, one can think of individual decisions as “more transparent” relative to the group or more informative regarding individual arbitrators’ views. Therefore, sole arbitrators are expected to be more affected by reputation concerns. However, the arbitration environment is more likely to generate “adverse reputation effects”, i.e. concerns for “bad” reputation as biased rather than expectations for improved

---

6To some extent, in FINRA arbitration, this is formally set by regulations, since only arbitrators that qualify as chairs can serve as sole arbitrators.

7Eliaz et al. (2006) present a theoretical model that aims to reconcile the conflicting empirical evidence.

8Following this idea, Bénabou (2013) presents a theoretical model of “collective denial and willful blindness in groups, organizations, and markets”.

9In a recent work, citets2016 harnessing present evidence for herding in decision making using data from a crowd-based earnings forecast platform. They also show that this has a negative impact on the accuracy of consensus forecasts.
reputation as experts, simply because the “truth” or the correct decision is never revealed. Therefore, the accuracy (or quality) of arbitrators’ decisions cannot be evaluated (Klement and Neeman, 2013). As a result, many hypothesized that arbitrators tend to simply “split the difference” between the parties to a dispute and avoid extreme decisions (Posner, 2004).

Relying on these insights, panels of arbitrators are expected to grant more extreme decisions when compared to individual arbitrators. From a normative perspective, it is not clear whether this effect of groups increases the accuracy of decisions. On the one hand, it allows biased arbitrators to “hide behind the group” and express their favoritism, hence reducing accuracy. On the other hand, impartial arbitrators will not attenuate their views where extreme decisions are warranted and increase accuracy. In the next section, I discuss the specific setting of FINRA arbitration and the economic forces and incentives that are expected to play a role in this setting.

3 Background on FINRA Arbitration

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. Its main operation is the formulation and enforcement of rules that are aimed to ensure fair and honest conduct of its members. In addition, FINRA runs the largest dispute resolution forum in the US securities industry, offering extensive arbitration and mediation services. Arbitration cases are divided into two types: customer cases, where the claimant is a customer and the respondent is a member firm, and industry cases, were both parties are either members or other entities who are involved in the industry, such as employees or suppliers.

In customer cases, which are the focus of this study, the arbitration process starts when an investor files a claim against a brokerage firm. Essentially all agreements between brokerage firms and their customers include a clause which obliges them to turn to FINRA arbitration in case of a dispute. Therefore, FINRA customer cases are considered to be mandatory arbitration proceedings, meaning that customers cannot choose to file their claim in court or in any other dispute resolution forum.

---

10 Generally, Morris (2001); 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 rather than on their consequences.

11 Although earlier papers suggest that this tendency is actually the result of parties’ positions in negotiations (Farber, 1981; Bazerman and Farber, 1985).

12 In some cases investors tried to turn to courts to discard them from the obligation to use arbitration proceedings. These appeals were mostly denied by the supreme court, even though courts and legislatures often expressed the opinion that mandatory arbitration clauses are unfair or might violate the basic principle of the justice system, that everyone is entitled to have their “day at court” (see Shearson/American Express Inc. v. McMahon (482 US 220 [1987]), Rodriguez de Quijas v. Shearson/American Express, Inc. (490 US 477 [1989]))
After filing the claim, the arbitrators’ selection process begins. The number of arbitrators appointed to a case depends on the amount of damages initially requested in the statement of claim. FINRA regulations set a threshold amount such that cases with a requested relief equal to or below it are heard by a sole arbitrator and cases above the threshold are heard by a panel of three arbitrators.\(^{13}\) However, parties may agree to stray from this rule.\(^{14}\) 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).\(^{15}\) Only public arbitrators that meet certain requirements of experience and education can serve as chairpersons.\(^{16}\) In cases where only one arbitrator is appointed, he must be a public arbitrator qualified to serve as chairperson. In cases with three arbitrators, at least two arbitrators are public arbitrators and at least one must be eligible for chairperson. 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.\(^{17}\)

FINRA uses the “veto-rank” method for the selection of arbitrators. Each party receives a list of randomly selected arbitrators and is asked to veto some and rank the others. The arbitrators with the highest combined scores are appointed. In cases where one arbitrator should be nominated, parties get one list of ten potential arbitrators. They can veto four out of ten and then they rank the remaining six. 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 total number of arbitrators assigned to the case.\(^{18}\)

A lot of effort is exerted to sustain the image of its arbitration proceedings and arbitrators as fair and objective, since they are often suspected of bias, especially in favor of the FINRA member firms. Nevertheless, both the FINRA forum and individual arbitrators were sometimes claimed to

---

\(^{13}\) It should be noted that the threshold was raised during the studied period and this change in regulations is used as the identification strategy (see section 4.2).

\(^{14}\) Before the change in regulations, cases that were assigned to a sole arbitrator, could be heard by a panel of three by request of the claimant (without requiring the consent of the respondent).

\(^{15}\) Non-public arbitrators are often referred to as “industry arbitrators”. The exact definitions are listed in rules 12100(p) & (u) of FINRA’s Code of Arbitration Procedure for Customer Disputes.

\(^{16}\) To qualify as chairperson an arbitrator must also complete FINRA’s chairperson training. The complete requirements for chairperson eligibility are listed in rule 12400(c) of FINRA’s Code of Arbitration Procedure for Customer Disputes.

\(^{17}\) This was the only option for the composition of the panel up to 2011 when the rules changed. Since January 2011, the claimant can choose to have an “All Public Panel”, i.e. appoint only public arbitrators to the panel. Between 2008 and 2011, there was a pilot program that offered the choice of an “All Public Panel” in selected cases. In my main sample only 46 cases were actually heard by an “All Public Panel”.

\(^{18}\) The only exception is in the case of an “All Public Panel” where the parties get only two lists of 10 and the selection process is slightly different. However, as mentioned above, such cases are negligible in our sample.
be “customer friendly”. In a recent study, Egan et al. (2018) find a negative association between arbitrators’ tendency to award high amounts of damages to customers and the likelihood of getting selected in future disputes. Their findings imply that parties, especially firms, consider past decisions of arbitrators to be a significant signal for potential bias.

The FINRA guidelines to arbitrators specifically address the question of how a decision should be made by a panel. The arbitrator’s guide offers the following protocol:

After each panel member has reviewed the case information, the chairperson should begin the discussion by asking the panel members to express their individual views of the case. All arbitrators must take part in deliberating the facts and issues, and their observations and opinions should be heard, acknowledged and considered.

In addition, rule 12410 of the FINRA Code states that all rulings and determinations by the panel shall be by a majority of the arbitrators. Note that while a decision requires at least two panel members to agree on liability, the reasons behind their determinations may differ. In addition, 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 the phenomena referred to as “dissent aversion” among judges (Epstein et al., 2011).

4 Empirical Analysis of FINRA Arbitration Awards

4.1 Data and Sample

The data was collected from FINRA arbitration awards, which are publicly available on-line. The complete database consists of 4,478 awards for customer cases and 2,291 awards for industry cases, that were filed between 2006 and 2011, and resolved by January 2013. For each of these cases, the record includes filing and award dates, the relief requested in the statement of claim and the awarded damages, in addition to the number of arbitrators and their identities.

19Scanning through securities blogs and law firm websites many statements that indicate that FINRA is biased in favor of its own members appear. One example is “There is a wide-spread perception that the arbitration system is stacked against customers. Some people have expressed concern that many arbitrators are reluctant to mete out large awards against big brokerage firms or to issue strong discovery rulings.” (http://www.sbplllplaw.com/an-outline-of-the-finan-rararbitration-process-for-customer-broker-disputes/). Others express exactly the opposite opinion, referring to FINRA as “customer-friendly” and suggesting that FINRA’s own members are interested in taking disputes with customers to other judicial forums (http://www.pepperlaw.com/publications/second-circuit-makes-it-easier-to-avoid-finra-arbitration-2015-04-02/). However, the latter view is much less prevalent.

20Out of 2575 panel decisions in the full sample only 78 cases record a dissenting arbitrator. In some of these cases, it is noted that the dissent only regards to a specific detail or a technical component of the award, e.g. expungement of records or distribution of costs between the parties.
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 Award Rate or AR).\textsuperscript{21} 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.\textsuperscript{22}

The focus of this study is on customer cases, where claimants are customers and respondents are firms. In this setting, parties belong to two distinct groups with opposing interests that repeatedly confront each other on similar matters. This increases the comparability of outcomes across cases. Moreover, only in this type of cases, litigants may worry about arbitrators being systematically biased in favor of customers or firms. Hence, reputation concerns may play an important role.

To further increase the comparability of cases, in the main analysis, the sample is restricted to cases where the amount of requested damages does not exceed $250,000.\textsuperscript{23} For this sub-sample, detailed information on parties and case characteristics was recorded, including the number of claimants and respondents and their representation, the causes of claim (or controversy type) and the types of securities which are the subject of the dispute.

Since the data is retrospective, in the sense that we only observe a case once the award is posted, censoring may affect the results. To avoid such a distortion, I use survival analysis and only consider cases where the award was issued within a limited period of time since the case was filed. This period is calculated as the 90th percentile of the distribution of case duration for each category of cases. Figure 1 presents the cumulative distribution of the duration for 3 groups of cases, divided by the relief requested. The cutoffs between groups are based on FINRA regulations regarding the number of arbitrators that are assigned to cases and are used to construct the treatment and control groups for the main analysis, as explained in subsection 4.2 below. To estimate the complete, uncensored distributions, only cases that were filed during 2006 and 2007 were included, assuming that for these two years, a 100% of cases that eventually result in a written award are observed in the sample.\textsuperscript{24} 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 end

\textsuperscript{21}cases where the requested damages are not specified or where no monetary relief is requested are excluded from the analysis.

\textsuperscript{22}In most cases the amount of punitive damages that were requested by the claimant is not specified, so it is impossible to use a consistent measure of total damages requested. Therefore, the relief requested includes only compensatory damages. In addition, as explained below, this is the amount that dictates the choice of forum size, which is the main interest in this study.

\textsuperscript{23}This choice is explained in further details in the next section.

\textsuperscript{24}More specifically, the assumption is that the number of cases that take longer than 5 years to complete is practically zero. The awards in the sample were collected until January 2013, allowing a maximal duration of 7 years for cases filed in the beginning of 2006, and 5 years to those filed very close to the end of 2007.
within 744 days. For the first control group, which is characterized by lower claims, the duration of proceedings tends to be shorter, and the opposite is true for the second comparison group. To portray how this limitation may affect the results, in section 4.4, a sensitivity analysis is conducted and boundaries are derived for the main estimates. In addition, the main analysis is repeated and results are confirmed using the full sample without any restriction and a more restricted sample that sets the limit at the 75th percentile for each group.

Last, it is important to note that aggregate statistics posted by FINRA suggest that approximately 80% of claims are terminated before the hearing stage, and hence no award is issued.\(^{25}\) Assuming that the selection of such cases is not correlated with the number of arbitrators, this does not pose a major concern for the purpose of the analysis, especially since in such cases the effect of the arbitrators on the outcome is negligible. Appendix figure A2 shows that the number of cases in the data is proportional to the total number of cases as reported in FINRA’s aggregate statistics, by filing year. In cases that are resolved in settlement at or after the hearing stage, an award is issued declaring that the parties settled, but the details of the settlement are usually kept secret.\(^{26}\) Obviously these outcomes should be included in the analysis, since settlements are not random and arbitrators’ behavior during the hearing could influence the probability of settlement. However, since the amount of settlement is unobserved, the award rate for these cases cannot be determined. Assuming that in a settlement agreement each party receives at least some of their initial claims, settlements are coded as ‘split the difference’ awards, namely award rates equal to 0.5.\(^{27}\) In section 5 this choice is further explored and a different method, predicting the AR based on observables, is used to confirm that this choice does not drive the main results.

### 4.2 Empirical Strategy

Comparing the awards granted by sole arbitrators to those of panels is not straightforward, since different types of cases are assigned to different numbers of arbitrators. Therefore, such a comparison might actually reflect differences in observable and unobservable characteristics of cases and litigants. In FINRA arbitration, the rules set a specific threshold for the assignment of three (rather than one) arbitrators to a case. This threshold is based on the relief requested by the claimant. Up to March 29th, 2009 the threshold for a single arbitrator was $50k and cases above this amount were assigned to panels of three. Figure 2 presents a ‘naive’ comparison of the distribution of

\(^{25}\)The reasons for closing a case at such a preliminary stage could be a direct settlement between the parties, bankruptcy of a critical party, forum denying, and others.

\(^{26}\)Such awards regularly decide on the issue of expungement, i.e. whether or not the claim will be erased from the member’s records, and on the fees that should be paid by the parties for the arbitration proceedings.

\(^{27}\)The average award rate for cases that did not resolve in settlement and where claims were not entirely dismissed is indeed 0.508.
Figure 1
Case Duration CDF (by Range of Relief Requested)
awards granted by panels of three versus sole arbitrators, for this period (which in the main analysis serves as the “pre-period”).

According to Figure 2, cases that were heard by sole arbitrators seem to have slightly more “extreme” awards, which grant plaintiffs with all or nothing, compared to cases with three arbitrators. However, this difference cannot be directly attributed to the different number of arbitrators, since these cases have many other observed and unobserved qualities which distinguish them (mainly, the amount of requested damages).

In 2009 the threshold for the assignment of three arbitrators was raised to $100k. 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 economic crisis evoked by the collapse of financial markets. This offers a unique opportunity to compare arbitration awards pre and post change, for cases with a relief requested between the old and the new threshold, i.e. between $50k and $100k. For this group of cases, if a claim was filed before Marh 30, 2009, the rules assigned a panel of three arbitrators to hear the case. Whereas after this date, only one arbitrator is assigned to each case in this same category. The nature and timing of the described change in FINRA regulations, can be used as a “natural experiment”, where the assignment of arbitral tribunal to cases in the affected group is as good as random.

Figure 3 presents the distribution of award rates before and after the change in regulations for the affected group of cases. This comparison clearly suggests that there was a sizable shift in the

---

28 See FINRA regulatory notice 09-13 “Threshold for Single Arbitrator Cases”.
29 Indeed 2009 stands out with an unusually large number of filed claims both in FINRA aggregate statistics and in my sample, as presented in Appendix figure A2.
distribution of awards and that the difference between panels’ and sole arbitrators’ decisions is actually opposite to the one suggested 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 where claims were denied entirely (or where the award rate was very close to zero) substantially decreased. A similar decrease is apparent for awards that grant claimants with the full amount of damages as 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.

These findings could support the theoretical prediction that groups tend to be more extreme. However, one should suspect that similar changes occur for all classes of cases and not just for cases in the affected group. This could be the result of a time trend or some other unobserved heterogeneity of the pre and post periods. Using a comparison group in a difference-in-differences (DD) design can refute this concern.

In this setting, there are two natural comparison groups: cases with a requested relief below $50k, which were assigned to sole arbitrators throughout the entire period and cases with a requested relief above $100k, which were assigned to panels of three arbitrators all along. To keep the different groups comparable and reduce variation within the comparison group, the entire sample (and hence the second comparison group) is limited to cases where the relief requested does not exceed $250k. The two comparison groups are used in the DD setting, while controlling for each category of cases separately to allow for variation in the average award rate across

---

30 The distribution of relief requested has a very long right tail. Approximately 40% of cases have a relief requested that is higher than $250k and the maximal relief requested for cases in the full sample is 1.25 billion dollars.
Therefore, the main specification for estimation is ($i$ is an index for a specific case):

$$Y_i = \alpha + \beta_0 \text{Treated}_i + \beta_1 \text{Post}_i + \beta_2 \text{Treated} \times \text{Post}_i + \gamma' X_i + \psi_i$$

Where $Y$ is the outcome of interest. The main outcomes that are considered are indicators for extreme and moderate awards. An extreme award is defined as an award rate that equals zero or one, meaning that one side to the dispute received precisely what they requested. A moderate award is an award rate close to 0.5, that can be interpreted as each side receiving approximately half of what they requested. More specifically, a moderate award is defined as an award rate between 0.4 and 0.6.

$\text{Treated}$ is an indicator for the case being in the affected range of relief requested (i.e. $50k$ to $100k$) and $\text{Post}$ indicates whether the case was filed after the change of rules came into effect. $\text{Treated} \times \text{Post}$ is the interaction between the two indicators and is the explanatory variable of interest in this setting. Therefore, $\beta_2$ is the difference in outcome that is attributed to the assignment of a case to a single arbitrator rather than to a panel of three arbitrators.

$X$ is a vector of controls, which can be divided into two groups. The first consists of variables that refer to the number of claimants and to whether or not the parties are represented by an attorney. The second includes a large set of dummy variables that describe the causes of claim and the type of securities involved in the dispute. A complete list of these variables is presented in Table 2 and the notes that follow. In addition, each specific category of cases within the comparison group is controlled for, including an indicator for cases where the relief requested is equal to or below $25,000 for which the “simplified arbitration” procedure applies.\textsuperscript{31}

Last, using data on arbitrators’ identity, the same specification is 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.

The main identifying assumption in this 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 due to the fact that the cutoffs that determine treatment status are somewhat arbitrary. While it is obvious that the rational for the rules that govern the number of arbitrators is that higher stakes call for more decision-makers, the specific choice of cutoff has no

\textsuperscript{31}Under this procedure, claimants can choose to only submit written pleadings and evidence or to have a hearing conducted by telephone conference call, in order to accelerate the proceedings.
specific reasoning. In addition, as noted above, the regulatory change was driven by the desire to alleviate caseload and therefore could be considered as exogenous to the outcomes of interest. It should be noted, that the quasi-arbitrary cutoff calls for consideration of 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.

A common way to test this assumption is to compare pre-change trends between treatment and control. In section 4.4 the main DD specifications are estimated using only the pre-period and a placebo treatment date. This estimation points to the absence of differential pre-trends for treatment and comparison groups, which further supports the validity of the empirical design.

Despite this, a major threat to identification is if litigants can somehow manipulate the rules. In this setting, the claimant actually determines the relief requested and can potentially alter this amount to affect the number of arbitrators. To assess this possibility, the distribution of the requested amounts is plotted separately for the pre and post periods and presented in Figure 4. A Kolmogorov-Smirnov test for equality of distributions is 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).

Additionally, in Table 4, I test the robustness of the main results to an alternative definition of the treatment and control group, where the threshold amounts are attributed to groups, as though manipulation actually occurred.
Table 1
Testing for Manipulation on Filing date

<table>
<thead>
<tr>
<th></th>
<th>Treatment (1)</th>
<th>Control (2)</th>
<th>Difference (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Proportion Filed Pre (Jan-Jun, 2009)</td>
<td>.472</td>
<td>.496</td>
<td>-.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.077)</td>
</tr>
<tr>
<td>(b) Proportion Filed Pre (Jan-Dec, 2009)</td>
<td>.248</td>
<td>.261</td>
<td>-.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.049)</td>
</tr>
</tbody>
</table>

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 of 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 for the difference. The sample is restricted to avoid censoring, as explained above in figure 1 and the text that follows.

One more concern is that parties can manipulate the number of arbitrators by changing the filing date (around the time that the new rule became effective). The change in rules was announced on February 2009, so the concern is that claimants either rushed to file their claims before the rule came into effect or delayed filing till after that date, hence actually selecting the number of arbitrators assigned to their case. If this was true, we would expect to see some irregularity in the number of cases filed by month around the time of the policy change. In fact the trends in number of new claims filed over the first 6 months of 2009 appear to be practically identical when the affected group of cases is compared to the control group. Table 1 presents the proportion of cases that were filed during the quarter that preceded the effective date of the new rule out of cases 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 found to be very small and statistically insignificant.

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

When considering manipulation, it should also be noted that parties can agree to change the size of the arbitral tribunal and moreover, prior to the change the consent of the respondent was not always required when the rule assigned a sole arbitrator to the case. This structure significantly reduces incentives to alter either the relief requested or the filing date in the interest of choosing panels instead of sole arbitrators or vice versa.
Figure 5
Claim Filing Frequency Pre & Post Change (by Range of Relief Requested)
Nonetheless, this possibility opens another important discussion. Allowing the litigants to agree on a different number of arbitrators than the regulations set means that compliance might be partial. Generally, compliance rates are very high in the sample. For the comparison group, more than 95% comply with the rules. For the treatment group, compliance is slightly lower but still high at 89%.\footnote{Litigants seem to comply more with the three arbitrators rule than with the sole arbitrator rule, especially in the treatment group (it seems reasonable to get lower levels of compliance to a new rule).}

In the main part of the paper, I compare cases by the expected number of arbitrators (according to the rules) to avoid bias caused by parties’ self selection of the tribunal size.\footnote{Conducting the same analysis dropping the cases where the actual tribunal is not the expected one, does not affect results significantly. However, since compliance cannot be considered random, such results could be biased and cannot be interpreted effectively.} Hence, results should be thought of as intention to treat effects (ITT), rather than actual treatment effects. Still, this is considered as the main specification because non-compliance rates are so low.

In addition, an Instrumental Variable analysis is used in order to derive the treatment effect on the compliers. An indicator for the rule change interacted with being in the treatment group serves as an instrument for forum size. This is a valid instrument since it is highly correlated with the the number of arbitrators actually assigned to a case and it affects the outcome only through the endogenous treatment variable (number of arbitrators). This strategy is discussed in more detail in section 4.5.

Going back to the actual regulatory change and assuming that there is no manipulation, one should still address the possibility that other changes occurred specifically in the treatment group over the same time periods. Such changes are expected to be portrayed in the composition of the treatment group 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, a DD specification as in equation (1) is estimated using each of the covariates as an outcome separately. The DD coefficient estimates for this set of regressions are presented in table 2. Out of the 39 characteristics that are tested only two changes are significant at conventional levels and one more is marginally significant. This is roughly what would be expected even if assignment to treatment was truly random. Nevertheless, in all the following estimations, the entire set of covariates is included. Table 2 also presents the mean of each covariate for the sample of cases that are used in the main analysis.

4.3 Results

Table 3 reports the estimated coefficients on the interaction term in the DD specification, with and without arbitrator fixed-effects, for the main outcomes of interest. The sample for the fixed-effect
Table 2
Descriptive Statistics & Stability of Group Composition

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>DD</th>
<th>Variable</th>
<th>Mean</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Claimants</td>
<td>1.379</td>
<td>-.13</td>
<td>Counter Claim</td>
<td>.02</td>
<td>.011</td>
</tr>
<tr>
<td>Claimant Represented</td>
<td>.681</td>
<td>-.03</td>
<td>Respondent Represented</td>
<td>.981</td>
<td>.035**</td>
</tr>
<tr>
<td>Number of Respondents</td>
<td>1.734</td>
<td>-.109</td>
<td>.02</td>
<td>.02</td>
<td>.011</td>
</tr>
<tr>
<td>Respondent Represented</td>
<td>.918</td>
<td>.012</td>
<td>.02</td>
<td>.02</td>
<td>.011</td>
</tr>
<tr>
<td>Third Party</td>
<td>.019</td>
<td>.012</td>
<td>.02</td>
<td>.02</td>
<td>.011</td>
</tr>
<tr>
<td>Security type -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government Bonds</td>
<td>.024</td>
<td>.005</td>
<td>Margin Calls</td>
<td>.025</td>
<td>.002</td>
</tr>
<tr>
<td>Corporate Bonds</td>
<td>.097</td>
<td>.004</td>
<td>Churning</td>
<td>.049</td>
<td>-.013</td>
</tr>
<tr>
<td>Funds</td>
<td>.154</td>
<td>.029</td>
<td>Excessive Trading</td>
<td>.137</td>
<td>.006</td>
</tr>
<tr>
<td>Real Estate</td>
<td>.117</td>
<td>-.066</td>
<td>Failure to Supervise</td>
<td>.401</td>
<td>.053</td>
</tr>
<tr>
<td>CMOs and MBOs</td>
<td>.006</td>
<td>.009</td>
<td>Negligence</td>
<td>.518</td>
<td>.009</td>
</tr>
<tr>
<td>Certificates of Deposit</td>
<td>.004</td>
<td>.012</td>
<td>Omission</td>
<td>.271</td>
<td>.047</td>
</tr>
<tr>
<td>Mutual Funds</td>
<td>.126</td>
<td>.113***</td>
<td>Breach of Contract</td>
<td>.384</td>
<td>.057</td>
</tr>
<tr>
<td>Options</td>
<td>.03</td>
<td>.003</td>
<td>Breach of Fiduciary Duty</td>
<td>.578</td>
<td>.092*</td>
</tr>
<tr>
<td>Limited Partnerships</td>
<td>.007</td>
<td>-.099</td>
<td>Unsuitability</td>
<td>.35</td>
<td>.006</td>
</tr>
<tr>
<td>Annuities</td>
<td>.091</td>
<td>.002</td>
<td>Misrepresentation</td>
<td>.412</td>
<td>.025</td>
</tr>
<tr>
<td>Preferred Stocks</td>
<td>.039</td>
<td>.024</td>
<td>Fraud</td>
<td>.286</td>
<td>.022</td>
</tr>
<tr>
<td>Auction Rate Securities</td>
<td>.019</td>
<td>-.018</td>
<td>Failure to execute</td>
<td>.111</td>
<td>.005</td>
</tr>
<tr>
<td>Other</td>
<td>.043</td>
<td>-.021</td>
<td>Concentration</td>
<td>.026</td>
<td>.021</td>
</tr>
<tr>
<td>Various Securities</td>
<td>.224</td>
<td>-.08</td>
<td>Promissory Estoppel</td>
<td>.011</td>
<td>-.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elderly Abuse</td>
<td>.008</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Respondeat Superior</td>
<td>.14</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Unjust Enrichment</td>
<td>.02</td>
<td>-.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Conversion</td>
<td>.016</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deceptive Practices</td>
<td>.03</td>
<td>.023</td>
</tr>
</tbody>
</table>

Notes: The table reports means for all observed characteristics and coefficient estimates for a difference-in-differences estimation comparing each characteristic pre and post treatment for the treatment versus comparison group (as specified in equation (1)). To avoid censoring the duration of cases is limited as explained hereinafter in figure 1 and the text that follows. Standard errors are are clustered at the arbitrator level and reported in parenthesis (** p<0.05, * p<0.1).
model is smaller since singletons are dropped. Columns (1)-(2), present the estimated change in “extreme” awards, namely award rates equal to zero or one, due to the switch from three arbitrators to a sole arbitrator. The estimates are negative and significant, and the magnitude of the effect remains similar within arbitrator. The interpretation of this result is that sole arbitrators are less likely to award either claimants or respondents with everything they claimed.

At the same time, a positive and significant change occurs for moderate or “split the difference” awards, where the award rate is between 0.4 and 0.6, as shown 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. Table A1 presents the same estimates using a full set of indicators for AR value ranges. There, it is clear that the only significant changes occur either for extreme values or in the middle of the range.

The magnitude of the effect is very similar with and without arbitrator fixed-effect and suggests 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 67%, this constitutes an approximately 30% decrease in extreme awards when a sole arbitrator is presiding.

In columns (5)-(8), the extreme outcomes are separated by the prevailing party - claimant or respondent. The impact is apparent on both sides, although in some of the specifications significance levels are marginal. The decrease in significance for the fixed-effect estimation is expected as the sample size is much smaller when singletons are dropped and due to the large number of fixed-effects (more than 400) relative to the sample size.

According to these estimates, extreme awards in favor of the respondent increase by approximately 10 percentage points. A similar decrease in percentage points is also found on the other extreme. However, comparing this change to initial levels reveals that while respondents experience a 17% decrease in extreme wins, claimants’ extreme wins are practically eliminated when sole arbitrators preside, with a decrease of close to a 100% of initial level.

If changes are driven by the reputation channel as proposed above, then the asymmetric effect might be explained by the fact that respondents (firms) are repeat players in arbitration and as such, enjoy an informational advantage over customers (Egan et al., 2018). Therefore, the reputational penalty for extreme decisions in favor of claimants is substantially higher than for extreme decisions in favor of respondents. These results are further discussed in section 5.
Table 3
The Effect of Sole Arbitrators on Award-Rates

<table>
<thead>
<tr>
<th></th>
<th>Extreme</th>
<th>Moderate</th>
<th>Inclination of Extreme Decisions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>TreatedXpost</td>
<td>-0.205***</td>
<td>-0.219**</td>
<td>0.195***</td>
</tr>
<tr>
<td></td>
<td>(0.0585)</td>
<td>(0.0928)</td>
<td>(0.0460)</td>
</tr>
<tr>
<td>Controls</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Arbitrator FE</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Observations</td>
<td>1,920</td>
<td>1,255</td>
<td>1,920</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports difference-in-differences estimates, for outcome variables that indicate the award rate (decided on by either one or three arbitrators). All specifications include controls for case characteristics (as listed in table 1) and a set of indicators for filing quarter. The sample is restricted to avoid censoring, as explained in figure 1 and the text that follows. Standard errors clustered at the arbitrator level in parentheses.

4.4 Robustness of Results

4.4.1 Sensitivity Analysis

The results presented so far are only based on a sample of approximately 90% of all cases, due to the retrospective nature of the data. As explained in the previous section, I only use awards that are given within a specific number of days from filing. 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, since from the time the last case was filed (time of the rule change) we have almost 4 years of data on awards. This allows us to consider the sample for the pre period to be the full sample of cases filed prior to the change. For the post period, I add artificial observations that follow the “worst case” scenario for the prediction of interest. More specifically, I increase the control group by 10% and all new observations are assumed to show the pattern predicted for the treatment group, i.e. all awards are exactly 0.5. The opposite is done for the treatment group, where half of the added observation have award rate equal to one and the other half are equal to zero. Row (b) of table 4 presents the estimated coefficient on the interaction term for the main outcomes of interest, when the main specification is estimated using this extended sample. When compared to the main results (presented again for convenience in row (a)), these should be thought of as the lower bound estimates. Clearly, the point estimates exhibit the same trends although smaller
magnitudes of change (as expected). Significance levels are maintained. Keeping in mind that this scenario is extreme and highly unlikely, this experiment actually proves that under more realistic scenarios, even if the missing cases do not follow the same pattern, results hold both qualitatively and quantitatively.

4.4.2 More Robustness Tests

Rows (c)-(f) in table 4 display an array of additional robustness tests, that reinforce the main findings.

First, in rows (c) and (d) I test whether the results are sensitive to changes in the duration limit (aimed at avoiding censoring effects). Row (c) presents the results for a sample without any restriction on case duration,\(^{34}\) while row (d) shows the estimates when a more stringent restriction is applied, i.e. the number of days at 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 FE specification, where the effect is smaller (but still a 14.5 percentage point decrease) and only marginally significant. However, sample size is very small for this specification (below 1000 observations).

Next, in row (e), I restrict the sample to exclude all cases filed during 2009, the year of change. The aim of this is twofold. First and foremost, avoiding the direct effects of the financial crisis. It is reasonable to assume that most claims that relate to the events of the crisis were filed during this year, especially since, as mentioned above, 2009 stands out with a very high number of claims relative to other years. Second, if there was some manipulation of file dates around the time of the rule change, it assures us that this does not drive the results. The estimated effects are actually larger and significant, although sample size is much smaller.

In row (f) the treatment and control groups are redefined to account for possible manipulation around the 50k and 100k cutoffs. In the main specification, cases with a relief requested equal to 50k are included in the control group both before and after the change. Assume that pre-change claimants tended to choose this amount to make sure that their case is heard by a sole arbitrator, even though they actually intended to request a slightly higher amount. Then actually in the pre period some of these cases should be in the over 50k group. However in the post period, this threshold is no longer relevant and there is no basis to believe any manipulation occurs. In accordance with this story, the 50k cases are moved from the control group into the treatment group in the pre period (but not in the post period). The same idea is applied to the 100k threshold, where the threshold cases are moved from the control group to the treatment group in the pre period.

\(^{34}\)The last filing year 2011 is dropped to avoid severe censoring.
## Table 4
Robustness & Placebo Tests

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

**Controls**

- √
- √
- √
- √

**Arbitrator FE**

- √
- √

**Notes:** The table reports difference-in-differences estimates, for the main outcome variables, using different samples and restrictions to test the robustness of the main results (presented for convenience in row (a)). In rows (b)-(d) I test the sensitivity of the main results to the sample restriction by case duration (aimed at avoiding censoring as explained above). In row (b) “fake” observations that follow the ‘worst case’ scenario are added to the post period, and hence the estimates present the lower bound estimates for an unrestricted sample. In row (c), the sample is not limited based on duration but the last filing year (2011) is dropped (to avoid severe censoring). In row (d) duration is limited at the 75th percentile of duration for each category of cases (by relief requested). Row (e) presents the results for the same specifications using a sample that excludes all cases filed during 2009, which was the year when the number of claims peaked due to the financial crisis. Specification (f) shifts the threshold between the groups of cases to account for potential manipulation around the cutoffs. Row (g) reports DD-placebo estimates using only cases from the pre-change period and setting a placebo change date at July 15th 2007 (splitting the pre period in half). All specifications include controls for case characteristics (as listed in table 1) and a set of indicators for filing quarter, except for the sensitivity analysis in row (b), where control variables were not generated for the “fake” observations. The sample is restricted to cases where the duration of proceedings is lower or equal to the 90th percentile of case duration in each group of cases, except where noted otherwise. Standard errors clustered at the arbitrator level in parentheses.

*** p<0.01, ** p<0.05, * p<0.1
period (but not in the post period). This re-definition does not change the main conclusions as the estimates remain practically identical for extreme awards and in column (3) for moderate awards. When arbitrator FE's are added, the estimated impact on moderate awards drops slightly below conventional significance levels.

Overall, the different robustness tests confirm the patterns reported in the main analysis.

4.4.3 Placebo Test

To further establish that results are not coincidental or simply follow a pre-existing trend, the last row of table 4 present the results of a placebo test. Row (g) uses the same treatment and control groups but only in the pre period. This period is divided into two equal sub-periods, setting August 15th 2007 to be the placebo-treatment date, so the placebo is on the “post” indicator. None of the effects are significant and point estimates do not resemble the main findings. These results are helpful in refuting concerns regarding any difference in pre-trends between treatment and control groups.\(^{35}\)

4.5 Instrumental Variable Estimation

All the preceding results are based on the expected number of arbitrators according to the regulations applied by FINRA at the filing date. However, as mentioned above, regulations allow parties to agree to change the number of arbitrators. Parties may wish to switch from a panel of three to a sole arbitrator to reduce costs of proceedings. They may also choose to switch based on their perception of which forum will better serve their interests. In practice, most parties seem to comply with the rules set by FINRA (as reported above, compliance rate is approximately 90 percent).

The lack of full compliance means that the results displayed so far should be thought of as the “Intention to Treat” effects. Namely, the portrayed change is 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 compliers, I next present the results of an estimation that uses the interaction between the rule change and the affected group as an Instrumental Variable for the actual number of arbitrators.

The proposed instrument satisfies the necessary conditions and hence estimators reported in Table 5 are consistent. We established above that the regulatory change can be viewed as exogenous to the outcomes of interest. It is reasonable to assess that the exclusion restriction holds since

\(^{35}\) Although only the results for one placebo-treatment date is reported in the table, this exercise was conducted for several different cutoffs.
Table 5
The Effect of Sole Arbitrator on Award Rates - IV Approach

<table>
<thead>
<tr>
<th></th>
<th>1st Stage</th>
<th></th>
<th>2SLS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Treated × Post</td>
<td>0.716***</td>
<td>0.706***</td>
<td>-0.283***</td>
<td>-0.304**</td>
<td>0.269***</td>
</tr>
<tr>
<td></td>
<td>(0.0347)</td>
<td>(0.0523)</td>
<td>(0.0804)</td>
<td>(0.131)</td>
<td>(0.0639)</td>
</tr>
<tr>
<td>Sole public arbitrator</td>
<td>-</td>
<td>-</td>
<td>-0.283***</td>
<td>-0.304**</td>
<td>0.269***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0804)</td>
<td>(0.131)</td>
<td>(0.0639)</td>
</tr>
<tr>
<td>Controls</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Arbitrator FE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,920</td>
<td>1,255</td>
<td>1,920</td>
<td>1,255</td>
<td>1,920</td>
</tr>
</tbody>
</table>

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 a set of indicators for filing quarter. The sample is restricted to avoid censoring, as explained in Figure 1 and the text that follows. Standard errors clustered at the arbitrator level in parentheses.

*** p < 0.01, ** p < 0.05, * p < 0.1

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. Columns (1)-(2) present the results for the first stage (with and without arbitrator FEs) and confirm that the instrument is highly correlated with a sole arbitrator presiding over the case. In addition, the F-statistic for the first stage is above 190.

Columns (3)-(6) present the 2SLS estimators for the main specifications and outcomes. As expected, these estimates are larger in absolute values compared to the DD estimators reported above in Table 3 and highly significant. Generally, it shows that the effect is larger for cases that ‘play by the rules’ and reinforces the conclusion that panels and sole arbitrators behave and decide differently.

5 Exploring the Underlying Mechanisms

The main results, as confirmed by the multiple robustness tests, reveal a tendency of panels to give more extreme awards when compared to singles. The strongest impact seems to be on customer wins, implying that sole arbitrators are more reluctant to rule against brokerage firms than the other way around.

Three alternative explanations can be offered to the observed differences between individual
and group decisions in this setting. First, litigants may have different preferences over specific attributes of arbitrators, that depend on whether the chosen arbitrator will serve as a sole arbitrator or as a member of a panel (or chair). For example, parties may put more weight on arbitrators’ experience when they choose a sole arbitrator. If these attributes are also correlated with the potential decision, then the results presented above remain un-biased estimators of the impact of the rule change, namely the difference in outcomes between a regime where individual arbitrators decide on cases and a regime where groups decide. However, they cannot be interpreted as differences between group and individual decisions. To explore this option, Table 3 also reports the results of an arbitrator fixed-effects estimation. Since only arbitrators that qualify as chairs may serve as sole arbitrators, the fixed-effect attributed to each panel of arbitrators is based on the identity of the panel-chair. In addition to the technical reason for this choice, it is reasonable to expect that the chair will have the greatest influence on the decision of the panel, in light of the extensive responsibilities and authorities assigned to panel-chairs by FINRA’s regulations.

It should be noted, that this specification dramatically decreases statistical power, since there are 456 unique arbitrator names in a sample of 1255 cases. The “within” estimators for the change in extreme, zero or one, awards are negative, implying that the same arbitrator renders less extreme awards as a sole arbitrator compared to the awards issued by the panel she chairs. These estimates are remarkably similar to the results without fixed effects, although not always significant at conventional levels (but very close). Hence, these results suggest that arbitrators change their decision patterns depending on the type of tribunal, and tend towards less extreme decisions when deciding as individuals (rather than within groups). Alternatively stated, these results rule out the “selection mechanism” suggested above.

The second mechanism that potentially drives the results is the group polarization phenomena, which implies in the context of this research, that panels of arbitrators will always tend to give zero or one awards, regardless of the type of dispute they are asked to decide on. In contrast, the third potential channel predicts that group polarization will only occur in the presence of reputation concerns. When arbitrators are penalized for extreme decisions by gaining a “bad” reputation as biased, sole arbitrators will be more cautious than arbitrators in panels of three. With group decisions, the individual opinions are at least partially obscured and hence it is harder to infer bias from such decisions.

In the main estimations, I only focus on customer cases, where the claimant is always the investor 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. Hence, it is reasonable to assume that reputation concerns play a role.

However, in other cases (“industry cases”), the parties to the dispute could be two firms or
Table 6
The Effect of Sole Arbitrator on Award Rates in Industry Cases

<table>
<thead>
<tr>
<th></th>
<th>Extreme (1)</th>
<th>Extreme (2)</th>
<th>Moderate (3)</th>
<th>Moderate (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TreatedXpost</td>
<td>0.0258</td>
<td>-0.115</td>
<td>0.00169</td>
<td>-0.00594</td>
</tr>
<tr>
<td></td>
<td>(0.0549)</td>
<td>(0.105)</td>
<td>(0.0299)</td>
<td>(0.0576)</td>
</tr>
<tr>
<td>Controls</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Arbitrator FE</td>
<td></td>
<td>√</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Observations</td>
<td>1,239</td>
<td>544</td>
<td>1,239</td>
<td>544</td>
</tr>
</tbody>
</table>

Notes: The table reports difference-in-differences estimates, for outcome variables that indicate the award rate (decided on by either one or three arbitrators), for a sample of FINRA industry cases. All specifications include a set of indicators for filing quarter. The sample is restricted to avoid censoring, using the same rules as for the main sample (as explained in figure 1 and the text that follows). Standard errors clustered at the arbitrator level in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

a firm and a related person (employee, supplier etc.) and no specific inference on arbitrators’ bias can be made based on their decisions. Table 6 presents the results of an estimation of the same DD model as above using a sample of industry cases. These 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. These results are not similar in any way to the estimates on the customer cases sample and none of the estimates is significant. No distinct pattern of moderation or polarization can be identified in the industry cases sample. These findings strongly support the proposed reputation mechanism as the most probable channel that explains the systematic differences between group and individual decisions. Nevertheless, no normative statement can be made on these differences, since it is not clear if the extreme decisions that are avoided by sole arbitrators are correct decisions or biased decisions.

Last, I return to the discussion of the settlements that appear in the sample. In table A2 the outcome that is considered first is an indicator for settlement. Results in columns (1)-(2) show a significant increase of 16 to 19 percentage points in the likelihood of settlements when panels of three are replaced by sole arbitrators. This result contradicts the well known prediction, that as costs of trial increase so does the probability of settlement (Priest & Klein (1984), Bebchuk (1984)). According to this, since panels of three impose higher costs on the parties when compared to a sole arbitrator, panels are expected to promote settlements. It also contradicts the findings of Marselli et al. (2015) who report that settlements are associated with the presence of three arbitrators rather than one. The explanation that they offer to the findings is that in such cases results are more expected and this fosters the successful bargaining between the parties. In the FINRA sample,
panels actually seem to be less expected as the variation of group decisions is obviously larger. Therefore, their explanation may actually fit both results.

At the same time, the findings regarding settlements seems to be in line with the reputation channel, if one believes that arbitrators can effectively influence the probability of settlement. Since settlements will not affect the arbitrator’s reputation, sole arbitrators will be more interested in settlements and therefore act upon it.

6 Conclusion

This study presents empirical evidence on the difference between decisions of sole arbitrators and panels of three arbitrators. The results portray a tendency of panels to make more extreme decisions, i.e. closer to zero or one in award rates. In addition, panels seem to be less inclined to grant ‘split-the-difference’ awards or to promote settlements.

The results are based on a research design that exploits a change in FINRA regulations regarding the threshold for assigning cases to panels rather than sole arbitrators, which offers a rare opportunity to compare group and individual decision-making. Using a difference-in-differences setting, allows to estimate the causal effect of “singles versus panels”. Interestingly, results contradict many theoretical predictions and other findings based on observational studies. Mainly, these results are surprising in light of the prevalent conjecture in the arbitration context that panels are more predictable, since arbitrators tend to balance each other’s bias or noisy signals by averaging out the opinions of the panel members.

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 and hence have the courage to make extreme and even controversial decisions. If the arbitrators derive these results based on their true evaluation of the facts of the case this result is desirable. However, such results may also be driven by biased arbitrators, who use the fact that they are part of a group to freely express their favoritism. Without the support of the panel, the same arbitrators are expected to make less biased decisions, to avoid damaging their reputation.

While the conclusions of this study can be extended to other contexts and settings where decisions could be made by either groups or individuals, this should be done with caution. Theoretically, it is expected that different mechanisms will 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 (especially in social psychology, management, law and political science), finding more opportunities to empirically compare groups and individuals outside labs is essential.
References


7 Appendix

Figure A1
Number of Cases by Filing Year

Notes: Figure (a) on the left hand side, presents the number of cases filed with FINRA arbitration every year, between 2006 and 2011. The source for these numbers is 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 the end of 2012 are counted, and this probably explains the relatively small number of cases in 2011.

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 frequency of number of cases filed with FINRA arbitration every year, between 2006 and 2011. The source for these numbers is 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 the end of 2012 are counted, and this probably explains the relatively small number of cases in 2011.
Table A1
The Effect of Number of Arbitrators on Award Rate Distribution

<table>
<thead>
<tr>
<th></th>
<th>(1) 0 ≤ AR &lt; .2</th>
<th>(2) .2 ≤ AR &lt; .4</th>
<th>(3) .4 ≤ AR &lt; .6</th>
<th>(4) .6 ≤ AR &lt; .8</th>
<th>(5) .8 ≤ AR &lt; 1</th>
<th>(6) AR ≥ 1</th>
<th>Avg. AR</th>
<th>FE</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) TreatedXpost</td>
<td>-0.123** (0.0583)</td>
<td>0.0217 (0.0323)</td>
<td>0.195*** (0.0460)</td>
<td>0.00503 (0.0267)</td>
<td>-0.000806 (0.0235)</td>
<td>-0.0980*** (0.0312)</td>
<td>-0.0139</td>
<td></td>
<td>1920</td>
</tr>
<tr>
<td>(b) TreatedXpost</td>
<td>-0.155* (0.0869)</td>
<td>0.0156 (0.0406)</td>
<td>0.182** (0.0711)</td>
<td>-0.00274 (0.0285)</td>
<td>0.0492 (0.0351)</td>
<td>-0.0885* (0.0535)</td>
<td>-0.0109</td>
<td>√</td>
<td>1255</td>
</tr>
</tbody>
</table>

Notes: Standard errors are clustered at the arbitrator level and reported in parenthesis (*** p<0.01, ** p<0.05, * p<0.1).
### Table A2
Evaluating Settlements

<table>
<thead>
<tr>
<th>Settlements Re-coded</th>
<th>Settled</th>
<th>Extreme</th>
<th>Moderate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>TreatedXpost</td>
<td>0.162*** (0.0396)</td>
<td>0.189*** (0.0622)</td>
<td>-0.205*** (0.0585)</td>
</tr>
<tr>
<td>Controls</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Arbitrator FE</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Observations</td>
<td>1,920</td>
<td>1,255</td>
<td>1,920</td>
</tr>
</tbody>
</table>

Notes: Standard errors clustered at the arbitrator level in parentheses.  
*** p<0.01, ** p<0.05, * p<0.1