DISCUSSION PAPER SERIES

No. 8724

ON THE SELECTION OF ARBITRATORS

Geoffroy de Clippel, Kfir Eliaz
and Brian Knight

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP8724.asp**

ISSN 0265-8003

ON THE SELECTION OF ARBITRATORS

Geoffroy de Clippel, Brown University
Kfir Eliaz, Brown University and CEPR
Brian Knight, Brown University

Discussion Paper No. 8724
December 2011

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 8724

December 2011

ABSTRACT

On the Selection of Arbitrators*

A key issue in arbitration, which resolves disputes among parties, involves the procedure for selecting an arbitrator. We take an implementation-theoretic approach and provide theoretical, empirical and experimental analyses of this problem. Our findings highlight the problems with current procedures and suggest that alternative procedures, which we propose, may be superior.

JEL Classification: C72, C92 and D02
Keywords: arbitration, implementation

Geoffroy de Clippel
Department of Economics
Brown University
64 Waterman St.
Providence, RI 02912
USA

Email: declippel@brown.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=170859

Kfir Eliaz
Department of Economics
Brown University
64 Waterman St
Providence, RI 02912
USA

Email: Kfir_Eliaz@brown.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=153104

Brian Knight
Department of Economics
Brown University
64 Waterman St
Providence, RI 02912
USA

Email: brian_knight@brown.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=174819

* We wish to thank Eli Zvuluni of Possible Worlds Inc. for programming the experiment, Melis Kartal and Mark Bernard for running the experiment, CESS at NYU and especially Caroline Madden for invaluable administrative help, Samuel Menco, Pantelis Solomon, Ee Cheng Ong and especially Neil Thakral for exceptional research support.

Submitted 11 December 2011

On the Selection of Arbitrators

Geoffroy de Clippel, Kfir Eliaz, and Brian Knight*

November 29, 2011

Abstract

A key issue in arbitration, which resolves disputes among parties, involves the procedure for selecting an arbitrator. We take an implementation-theoretic approach and provide theoretical, empirical and experimental analyses of this problem. Our findings highlight the problems with current procedures and suggest that alternative procedures, which we propose, may be superior.

1 Introduction

Implementation theory studies the design of institutions and procedures for collective decision-making such that in equilibrium, the participants select the outcomes that are deemed "desirable", given the participants' preferences. The criterion for desirability varies across situations and is represented by a "social choice rule" (SCR), which maps the participants' preferences to subsets of feasible outcomes (e.g., the set of efficient outcomes, the set of "fair" outcomes, according to some notion of fairness). Many prevalent procedures for reaching collective decisions may be interpreted as attempts to construct mechanisms that implement some SCR by aggregating the participants' preferences in a

*Brown University, Department of Economics, Providence, Rhode Island - declippel@brown.edu, kfir_eliaz@brown.edu, brian_knight@brown.edu. We wish to thank Eli Zvuluni of Possible Worlds Inc. for programming the experiment, Melis Kartal and Mark Bernard for running the experiment, CESS at NYU and especially Caroline Madden for invaluable administrative help, Samuel Mencoff, Pantelis Solomon, Ee Cheng Ong and especially Neil Thakral for exceptional research support.

way that is deemed desirable. Examples include voting rules in hiring decisions, legal procedures for selecting a jury for a trial, systems for electing public officials and schemes for selecting arbitrators. A natural question that arises is whether prevalent mechanisms such as these indeed implement the intended SCR (and if so, under what conditions)? In addition, one may ask if these SCRs are implementable by alternative mechanisms? If not, are there variants of the original SCR that are implementable? How do the proposed mechanisms, which may work in theory, perform in practice?

This paper tries to address these questions in the context of assigning disputing parties to an arbitrator. We present theoretical, empirical and experimental findings that (*i*) point to several key weaknesses of the prevalent mechanisms, and (*ii*) suggest that alternative schemes, which we propose, may be preferred.

There are several advantages for focusing on the selection of arbitrators. First, arbitration is the most common procedure for resolving disputes without resorting to the costly process involved in courtroom litigation. In arbitration, the parties to a dispute agree to bring the case to a third party, the arbitrator, whose final decision is legally binding. Arbitration is used in a wide variety of areas including international trade agreements, commercial disputes in a vast array of sectors, civil disputes, labor disputes and sports. A key feature of arbitration is that the parties themselves play a role in appointing the arbitrator. Consequently, there are many agencies, both public and private, that offer their service in assigning arbitrators to cases. This is usually done by providing the parties to a dispute with a detailed list of potential arbitrators and employing some procedure for selecting an arbitrator from this list. Since the arbitrator's decision is final, the quality of the arbitration process critically depends on the properties of the mechanism used for selecting the arbitrator.

Second, most disputes that reach arbitration occur between parties that have a long-term relationship (e.g., unions and managements). In addition, the arbitration agencies provide the two parties with detailed resumés of the potential arbitrators. This suggests that, as a starting point, it is natural to assume that the parties' preferences (i.e., their ordinal rankings of the available arbitrators) is common knowledge between the parties but are not known to outsiders. This simplifies the analysis by allowing us to focus on implementation under complete information.[1]

In light of this, it is important to note that, perhaps contrary to simple intuition, there are reasons that parties' preferences over potential arbitrators may not be completely opposed. While some arbitrators may favor one party at the expense of the other, it is also the case that different arbitrators charge different fees, rule within a different length of time, and vary in their ability to fully understand the case. Indeed, in Section 4 our empirical analysis of the most common mechanism documents a substantial degree of overlap in parties' actions, which according to our theory, implies overlap in preferences.

Finally, the problem of selecting an arbitrator is a relatively simple and clean example of choosing a public good in an environment with symmetric information. There are only two participants, in most cases they face each other only once (or at least very infrequently) in arbitration, there are no monetary transfers between them at the point of selecting an arbitrator, the set of available objects is usually small (no more than seven in most cases) and only a single option is chosen (in contrast to jury selection where a subset if chosen).

The most common procedure for assigning arbitrators, the veto-rank scheme, is a simultaneous-move mechanism. The two parties receive a list with an odd number n of people who are accredited and available to rule the case. Both parties then have to select $\frac{n-1}{2}$ names out of this list, and rank the remaining $\frac{n+1}{2}$ candidates. The agency then compiles these two reports by first crossing names that have been selected by at least one of the two parties. Selecting a candidate's name thus amounts to placing a veto against her or him. Next, if multiple names remain (which happens whenever both parties select a same candidate at least once), then the agency picks the candidate that minimizes the sum of the ranks submitted by both parties, with ties resolved via a lottery.

[1] In his comprehensive survey of implementation theory under complete information, John Moore (1992) explicitly states that this informational environment is best suited to study bilateral situations where the participants had a long term relationship prior to participating in the mechanism.

When participants report their preferences truthfully (i.e., they veto the $\frac{n-1}{2}$ least preferred and rank the remaining ones according to their preferences), the appointed arbitrator will Pareto dominate both parties' median choices and be "sum-efficient" in the classical utilitarian tradition.

Truthful reporting, however, may not be a Nash equilibrium and non-truthful Nash equilibria may also exist. Empirical evidence from real arbitration cases (obtained from the New Jersey Public Employment Relations Commission (PERC) from 1985 to 1996) hints at non-truthful, strategic behavior (see Section 4). Strategic, non-truthful behavior is also exhibited by a significant proportion of participants in a laboratory experiment of the rank-veto scheme (see Section 6). If such behavior occurs in a rather abstract environment with small stakes and rather inexperienced subjects, then a fortiori should it happen in real-life conditions with larger stakes and professional "players" (most often lawyers).

Is there a way to design a procedure for which equilibrium outcomes coincide with the mapping from ordinal preferences to outcomes (i.e., the SCR) or alternatively, from von-Neumann-Morgenstern utility functions to lotteries over outcomes (what we call a random social choice function or RSCF) induced by implementing truthful strategies in the veto-rank procedure?[2] The answer turns out to be negative: we show that there is no simultaneous or purely sequential mechanism that guarantees that the selected arbitrator would be sum-efficient and satisfy the "minimum acceptance level" in the sense of dominating both parties' median choices. More generally, we show that no simultaneous mechanism can guarantee (in equilibrium) standard Pareto efficiency and a minimum acceptance level. We then propose alternative sequential procedures that do guarantee these two properties.

To the best of our knowledge, the only sequential procedure for selecting arbitrators in practice is the alternate strikes scheme (AS) (for a detailed - but non-exhaustive - list of agencies and the procedure they use, see Appendix 3). As the name indicates, both parties alternatively submit a name to remove from the list of potential arbitrators. The person selected to rule the case

[2]In Section 3 we present formal definitions of these concepts.

is the last to remain after all other names have been crossed. The subgame perfect equilibrium outcomes of this procedure define a rather natural iterative veto SCR (or RSCF when one takes into account the randomization over who moves first). Though the subgame perfect equilibrium strategies may be more complex, their associated outcome can indeed be derived through the iterated elimination of those names that are ranked worse by at least of the two parties among those candidates that remain available. It satisfies the two requirements of Pareto efficiency and Pareto dominance with respect to the parties' median choices.

We analyze two additional sequential procedures, which as far as we know, are not used in practice. In Voting by Alternating Offers and Vetoes (VAOV) players take turns in proposing arbitrators from a given list. When a proposed arbitrator is rejected by the other party, that arbitrator is removed from the list and the rejecting party then proposes a name from the remaining list. The procedure continues until a proposal is accepted or only one name remains (which is then selected). VAOV implements the Pareto efficient SCR/RSCF derived from the maximization of a simple natural social welfare function, where the arbitrator is picked among those candidates (at most two) that minimize the maximal rank over the two parties' preference orderings. It is thus the egalitarian alternative of the utilitarian criterion that was part of the RSCF derived from the rank-veto procedure played truthfully. The SCR/RSCF implemented by the VAOV is also clearly Pareto efficient.

The fourth and last procedure we investigate is motivated by the following observations: (*i*) sequential procedures are preferred to simultaneous-move games to avoid miscoordination, (*ii*) an extensive-form game should have very few rounds for participants to perform backward induction correctly (see e.g. Binmore et al, 2002), and (*iii*) the procedure should deliver outcomes that are Pareto efficient and Pareto dominate the parties' median choices. Proposition 6 establishes that there exists a unique two-round extensive form of perfect information whose subgame perfect equilibrium outcome satisfies (*iii*): a first party starts the game by selecting $\frac{n+1}{2}$ candidates, and the second party then selects the arbitrator out of that shortlist. We therefore refer to this procedure as shortlisting. Note that it may be viewed as the outcome of mechanism-design under a *behavioral constraint*, namely that backward induction cannot be performed for many rounds.

All four procedures are tested in a controlled lab experiment. We first show that a large fraction of the observed behavior is strategic and is consistent with either equilibrium play or a variant of the k-level reasoning (see Crawford et al. 2010). In addition, most participants in the sequential treatments exhibit a minimal form of backward induction. Interestingly, our data suggest that social preferences may play a role in the alternate strikes and shortlisting procedures (though such preferences may have a smaller impact between actual disputing parties in real arbitration cases). Finally, we compare the four procedures in terms of their performance on the two main criteria of efficiency (both Pareto and the utilitarian sum) and Pareto dominance over the parties' median choices. Our data suggests that the two procedures that are not used in practice – VAOV and shortlisting – perform better on both dimensions than the two procedures that are actually used in practice.

The paper unfolds as follows. After discussing the related literature in the next section, definitions and theoretical results are presented in Section 3. The empirical argument supporting the intuition that parties do not systematically report their preference truthfully is included in Section 4. The experimental design is presented in Section 5, while the data and its analysis are presented in the subsequent section. The concluding section summarizes our findings, and briefly discusses implementation in dominant strategies.

2 Literature Review

The most closely related paper is Bloom and Cavanagh (1986a), who analyze the selection of arbitrators using data on arbitration cases from the New Jersey Public Employment Relations Commission (PERC) during 1980. Data are based upon the simultaneous veto-rank scheme described in the Introduction (with $n = 7$). Their analysis first examines the degree of overlap between rankings in order to shed light on the similarity of preferences. They show some, but not complete, overlap in rankings, and, under the assumption of sincere rankings, conclude that there is some, but not complete, overlap in preferences. Their second analysis uses rankings and characteristics of arbitrators to measure the degree to which certain characteristics are valued by the different parties. They find, for example, that employers rank economists more highly than unions do. Under an assumption of sincere rankings, one can conclude that employers have a relative taste for economists and that unions have a distaste for economists.

The most closely related analysis focuses on a test of strategic play. In particular, they fit a model using only information from first choices and another using all of the rankings. Under the assumption that strategic players always rank their most preferred alternative first but strategize on the remaining rankings, the model using all rankings should generate parameter estimates similar to those using only the first choice data. The authors show that the parameters are indeed similar and conclude that there is no evidence of strategic play. A key limitation of this test involves the breakdown of the assumption that strategic players always rank their most preferred alternative first. It is straightforward to generate counter-examples to this: if the union vetoes the first choice of the employer, the employer may choose to not rank their most preferred alternative first as this is "wasting" the first ranking. Indeed, we present evidence from one of our experiments below documenting that a substantial fraction of players do not rank first their most preferred alternative when it is not viable, in the sense that this option is ranked last by the opponent.

In an unpublished working paper, Bloom and Cavanagh (1986b) discuss some theoretical properties of the rank-veto and AS mechanisms. They show that the former has non-truthful and inefficient equilibria, while the equilibria of the latter are all efficient. They also that if the parties held uniform priors over all the possible strict rankings of arbitrators, then being truthful is an efficient Bayesian Nash equilibrium in both mechanisms.[3] Our focus, however,

[3]One complication that arises when analyzing Bayesian Nash equilibria, especially in the rank-veto game, is that one needs to make assumptions about each player's belief about his

is on the implementation-theoretic view of arbitrator selection. In particular, we show that a large class of SCRs with appealing properties is impossible to implement, while other alternative SCRs are implementable by "natural" mechanisms.

Klement and Neeman (2011) consider the problem of how the selection procedure may affect an arbitrator's incentive to build a reputation of being neutral in order to get selected in the future. Their model of the selection procedure is different than ours: arbitrators are randomly drawn from a pool and assigned to the two parties until both simultaneously decide to accept one of them. The authors examine the effect of providing parties with information on the arbitrators' past decisions.

More generally, this present paper is related to a literature on matching, where economists have identified market failures and proposed new mechanisms that solve these failures. Several of these mechanisms, similarly to the veto-rank scheme used in selecting arbitrators, involve participants submitting rank-ordered preferences. Examples include mechanisms for matching residents to hospitals and students to elementary schools (see Roth (1984, 2007) and Abdulkadiroglu, Pathak, and Roth (2005a and 2005b)). This literature has focused on implementing strategy-proof mechanisms using variants of the Gale-Shapley deferred acceptance algorithm or the top-trading cycle mechanism. In the context of the selection of arbitrators, there is no deterministic mechanism that is strategy-proof and instead we propose and analyze alternative sequential mechanisms.

Given our focus on whether participant ranks and vetoes are sincere or strategic, this paper is also related to a literature on strategic voting, which can take many forms. In an experimental setting with three candidates and plurality rule, Forsythe, Myerson, Rietz, and Weber (1993 and 1996) find substantial evidence that voters are strategic in the sense of not voting for their most preferred candidate when this candidate has little chance of winning. Focusing on the case of bundled elections, Degan and Merlo (2007) find little evidence that voters are strategic in the sense that they might account for the

---

opponent's preferences over *lotteries.* This concern, however, is not discussed in their paper.

fact that policy outcomes may depend upon both the Congress and the President. In a model with incomplete information, Kawai and Watanabe (2010) estimate that a large fraction of voters in Japanese elections are strategic in the sense of conditioning on the state of the world where they are pivotal.

3 Theory

Let A be the finite set of $n \geq 4$ arbitrators that are proposed by the agency to rule the case. Let $i = 1, 2$ be the two parties involved. It is not implausible in our problem to assume that the parties' ordinal preferences are commonly known among them. Let $\mathcal{P}$ be the set of strict preference relations $\succ$ on A. A *social choice rule* (SCR) is a correspondence $f : \mathcal{P} \times \mathcal{P} \to A$, with the property that $f(\succ_1, \succ_2)$ is a non-empty subset of A, for each $(\succ_1, \succ_2) \in \mathcal{P} \times \mathcal{P}$. The SCR f is *partially implementable* if there exists a *mechanism* (S_1, S_2, μ), where S_i is i's strategy set and $\mu : S_1 \times S_2 \to A$ is the outcome function, such that, for each $(\succ_1, \succ_2) \in \mathcal{P} \times \mathcal{P}$, the set of pure-strategy Nash equilibrium outcomes associated to the strategic-form game $(S_1, S_2, \mu, \succ_1, \succ_2)$ is non-empty and a subset of $f(\succ_1, \succ_2)$.

Notice that the veto-rank procedure discussed in the Introduction does not qualify as a mechanism in this sense, because the outcome function delivers a lottery in some circumstances. Considering lotteries, and thinking about how parties behave when facing such uncertainty, forces us to consider risk preferences. Let $\mathcal{U}$ be the set of strict Bernoulli functions (the defining ingredient of von Neumann-Morgenstern preferences). A typical element u of $\mathcal{U}$ is thus simply a function $u : A \to \mathbb{R}$, with $u(a) \neq u(a')$ whenever $a \neq a'$, and preferences between lotteries over A are derived by computing expected utility with respect to u. It is less plausible to think that there is complete information regarding these Bernoulli functions, but our analysis is robust against that assumption in that our sole objective when considering lotteries is to show that strong negative results hold *even if* there was complete information in that regard. A *random social choice function* (RSCF) is a function $\psi : \mathcal{U} \times \mathcal{U} \to \Delta(A)$ that associates a lottery to each pair of strict Bernoulli functions. The RSCF

ψ is *implementable* if there exists a *random mechanism* (S_1, S_2, μ), where S_i is i's strategy set and $\mu : S_1 \times S_2 \to \Delta(A)$ is the outcome function, such that, for each $(u_1, u_2) \in \mathcal{U} \times \mathcal{U}$, any pure-strategy Nash equilibrium outcomes associated to the strategic-form game $(S_1, S_2, \mu, u_1, u_2)$ coincides with $\psi(u_1, u_2)$.

The veto-rank procedure is an example of random mechanism. Varying the number of options each party can veto leads to a large class of mechanisms. Given an integer k between 0 and $\frac{n}{2} - 1$, both parties $(i = 1, 2)$ simultaneously choose a pair (V_i, s_i), where V_i is a set of vetoed options that contains k elements from A, and s_i is a scoring rule that assigns to every element in $A \backslash V_i$ an integer from zero to $n - k - 1$ such that no two elements are assigned the same score. The outcome is determined as follows. If $A \backslash (V_1 \cup V_2)$ is a singleton, then this arbitrator is chosen. Otherwise, an element in $A \backslash (V_1 \cup V_2)$, is selected by maximizing the sum of scores, $s_1(\cdot) + s_2(\cdot)$, with ties being broken via a uniform lottery.

For each $a \in A$ and each $u \in \mathcal{U}$, let $\sigma(a, u) = \#\{a' \in A | u(a') < u(a)\}$. The veto-rank procedure is played truthfully if, for each $(u_1, u_2) \in \mathcal{U} \times \mathcal{U}$ and both $i = 1, 2$, the set V_i contains the k worst elements according to u_i, and $s_i(a) = \sigma(a, u_i) - k$, for each element $a \in A \backslash V_i$. This generates natural RSCFs, parametrized by an integer α between 0 and $\frac{n}{2} - 1$, the "minimum acceptance level". For each $(u_1, u_2) \in \mathcal{U} \times \mathcal{U}$, let $\psi_\alpha(u_1, u_2)$ be the uniform lottery defined over

$$\arg\max_{a \in X_\alpha} (\sigma(a, u_1) + \sigma(a, u_2)),$$

where

$$X_\alpha(u_1, u_2) = \{a \in A | \sigma(a, u_i) \geq \alpha, \text{ for } i = 1, 2\}.$$

The support of ψ_α also defines a natural SCR: for each $(\succ_1, \succ_2)$,

$$f_\alpha(\succ_1, \succ_2) := \mathit{support}(\psi_\alpha(u_1, u_2)),$$

where u_i is any[4] strict Bernoulli function that is consistent with $\succ_i$ over A. Notice that f_0 coincides with the classical Borda rule.

[4]Notice indeed that ψ_α varies only with the ordinal information encoded in the Bernoulli functions.

Truthtelling, however, is *not* always a Nash equilibrium in the veto-rank procedure. For example, suppose $A = \{a, b, c, d, e\}$ and the players have Bernoulli functions (u_1, u_2) that generates the following rankings on A: $a \succ_1 b \succ_1 c \succ_1 d \succ_1 e$ and $b \succ_2 a \succ_2 c \succ_2 d \succ_2 e$. Note that reporting truthfully (i.e., vetoing d and e and giving a score of 2 to the top ranked element, a score of 1 to the second-best element and a score of 0 to the remaining element) is not a Nash equilibrium of the normal-form game with $k = 2$. If players followed this naïve strategy, they would end up in a tie, where either a or b is randomly chosen. If, on the other hand, player 1 would veto b instead of d (since player 2 vetoes d and e anyway), then a would be chosen uniquely, which he prefers. Note that there also exist inefficient Nash equilibria in which players use weakly dominated strategies and both delete their top two options.

While truthful reporting may not be a Nash equilibrium for some profiles, the Nash equilibrium outcome may still be "desirable" in the sense of being an element of the SCR, given the true preference profile. Observe that c was an equilibrium outcome in the previous example, with both parties vetoing a and b, which contradicts $\psi_2(u_1, u_2)$ and does not belong to $f_2(\succ_1, \succ_2)$. On the other hand, one might argue that this Nash equilibrium is not likely to emerge since it involves dominated strategies. There are Bernoulli functions (u_1, u_2) and associated preference profiles $(\succ_1, \succ_2)$ for which reasonable Nash equilibrium outcomes do *not* belong $f_\alpha(\succ_1, \succ_2)$, and hence a fortiori cannot coincide with $\psi_\alpha(u_1, u_2)$. Consider again the case $k = 2$, and Bernoulli functions (u_1, u_2) generating the rankings $a \succ_1 b \succ_1 c \succ_1 d \succ_1 e$ and $e \succ_2 c \succ_2 a \succ_2 b \succ_2 d$. Then $f_2(\succ_1, \succ_2) = \{a\}$ and $\psi_2(u_1, u_2) = a$. However, there exists a (undominated) Nash equilibrium in which player 2 chooses $V_2 = \{a, b\}$ and s_2 such that $s_2(e) = 2$, $s_2(c) = 1$ and $s_3(d) = 0$, while player 1 chooses $V_1 = \{d, e\}$ and s_1 such that $s_1(a) = 2$, $s_1(b) = 1$ and $s_1(c) = 0$. The outcome of this equilibrium is c.

The two previous paragraphs raise the questions of whether there exists another normal-form mechanism that implements the RSCF ψ_α, or that partially implements the SCR f_α. The next proposition shows that the answer to both questions is negative.

Proposition 1 *Let α be any integer between* 0 *and* $\frac{n}{2} - 1$. *Then ψ_α is not implementable, and f_α is not partially implementable.*

Proof: Hurwicz and Schmeidler (1978) proved that any SCR that is Pareto efficient (usual definition - see below) and partially implementable must be dictatorial. Since f_α is Pareto efficient and non-dictatorial, it is not partially implementable. We now focus on the implementability of ψ_α. Suppose, to the contrary of what we want to prove, that ψ_α is implementable. Let a, b, c, d be four elements of A, and let $(u_1, u_2) \in \mathcal{U} \times \mathcal{U}$ be such that $u_1(a) > u_1(b) > u_1(c) > u_1(d) > u_1(x)$, $u_2(c) > u_2(d) > u_2(a) > u_2(b) > u_2(x)$, for each $x \in A \setminus \{a, b, c, d\}$, and $u_1(b) < (u_1(a) + u_1(c))/2$. Notice that $\psi_\alpha(u_1, u_2)$ picks a or c with equal probabilities. By Maskin monotonicity, $\psi_\alpha(u_1', u_2) = \psi_\alpha(u_1, u_2)$, where u_1' coincides with u_1, except that $u_1'(b) \in (u_1(d), u_1(c))$. Yet this is impossible since $\psi_\alpha(u_1', u_2)$ picks c for sure. This shows that ψ_α is not implementable. ■

We believe that a fundamental reason why agencies would like to implement $f_{\frac{n-1}{2}}$ with an odd number n of arbitrators, is that all the outcomes that emerge with positive probabilities satisfy the following two properties. A RSCF ψ is *Pareto efficient* if, for each (u_1, u_2) and each x in the support of $\psi(u_1, u_2)$, it is impossible to find $a \in A$ such that $u_i(a) > u_i(x)$ for both $i \in \{1, 2\}$. It passes the *minimal satisfaction test* if $\sigma(x, u_i) \geq \frac{n-1}{2}$ for each $i \in \{1, 2\}$, each $u \in \mathcal{U} \times \mathcal{U}$, and each x in the support of $\psi(u_1, u_2)$. Similar definitions also apply to SCRs. Note that both $f_{\frac{n-1}{2}}$ and $\psi_{\frac{n-1}{2}}$ are Pareto efficient and pass the minimal satisfaction test when n is odd. Unfortunately, they are not implementable. Yet one should wonder whether there might be other RSCFs and SCRs that are. The next propositions show that this is not the case.

Proposition 2 *Suppose that n is odd.*[5] *There is no SCR that is partially implementable, Pareto efficient, and that passes the minimal satisfaction test. There is no RSCF that is implementable, Pareto efficient, and that passes the minimal satisfaction test.*

[5] A comparable result can be stated and proved when there is an even number of arbitrators, but we focus on the case of an odd number since this is the scenario favored by agencies.

Proof: Hurwicz and Schmeidler (1978) proved that any SCR that is Pareto efficient and partially implementable must be dictatorial. Any such SCR will thus fail the minimal satisfaction test. We now pay attention to RSCFs. The proof is made for the case where A contains five elements - $A = \{a, b, c, d, e\}$ - but can easily be extended to any A with an odd number of elements. Consider (u_1, u_2) such that $u_1(a) > u_1(b) > u_1(c) > u_1(d) > u_1(e)$, and u_2 is completely opposite. If ψ passes the minimal satisfaction test, then $\psi(u_1, u_2)$ yields c with certainty. Maskin Monotonicity implies that $\psi(u_1', u_2')$ also yields c with certainty, where $u_1'(c) > u_1'(e) > u_1'(a) > u_1'(b) > u_1'(d)$ and $u_2'(e) > u_2'(c) > u_2'(a) > u_2'(b) > u_2'(d)$. Consider (u_1'', u_2'') such that $u_1''(c) > u_1''(a) > u_1''(e) > u_1''(b) > u_1''(d)$, and u_2'' is completely opposite. If ψ passes the minimal satisfaction test, then $\psi(u_1'', u_2'')$ yields e with certainty. Maskin monotonicity then implies that $\psi(u_1', u_2')$ also yields e with certainty, a contradiction. ■

The previous result shows that one must consider extensive-form mechanisms instead of simultaneous-move game forms if one wants to systematically derive desirable outcomes. Because our goal is to consider mechanisms that are potentially applicable, we focus on finite extensive-form mechanisms of perfect information, which are thus solvable by backward induction. In addition, we focus on a strong notion of implementability that combines ideas of partial implementability of SCRs and implementability of RSCFs. A SCR f is *fully implementable by backward induction* if there exists a two-player extensive-form mechanism of perfect information such that, for each $(\succ_1, \succ_2) \in \mathcal{P} \times \mathcal{P}$, $f(\succ_1, \succ_2)$ coincides with the union of the two subgame-perfect equilibrium outcomes[6] associated with the two extensive-form games obtained when assigning either the first or the second party to the role of the first player. A fully implementable SCR naturally leads to an RSCF by picking an element of the SCR via the throw of a fair coin. This associated RSCF is clearly implementable by backward induction, via the extensive-form where chance decides in a first move who will assume the role of the first player.

[6] Preferences being strict, backward induction always leads to a unique outcome in each extensive-form game of perfect information.

Implementability, efficiency and the minimal satisfaction test become compatible when considering backward induction in this larger class of mechanisms. Before discussing some procedures with these properties, we show that f_α is not fully implementable by backward induction. In fact we will even show that there is no single-valued selection of f_α that is implementable by backward induction. In other words, one is forced to consider alternative (desirable) SCRs.

Proposition 3 *Let* α *be any integer between* 0 *and* $\frac{n}{2}-1$*. Then* f_α *is not fully implementable by backward induction.*

Proof: We will show the stronger result that there is no single-valued selection of f_α that is implementable by backward induction. Suppose, to the contrary of what we want to prove, that there exists an extensive-form mechanism that leads to backward induction outcomes that systematically fall within f_α. Let a, b, c, d be four elements of A. Consider the following pair $(\succ_1, \succ_2)$ of orderings: $a \succ_1 b \succ_1 c \succ_1 d \succ_1 x$ and $d \succ_2 c \succ_2 a \succ_2 b \succ_2 x$, for each $x \in A \setminus \{a, b, c, d\}$. The backward induction outcome computed for this pair of preferences must be a, since this is the only element in $f_\alpha(\succ_1, \succ_2)$. Let now $\succ_2'$ be the same preference ordering as $\succ_2$, except that the relative ranking of c and d is reversed. We now prove that a must be the backward induction outcome of the mechanism when computed for $(\succ_1, \succ_2')$. For each decision node ν, let $\mathcal{O}(\nu, \succ_1, \succ_2)$ be the set of arbitrators that the party in charge at ν can generate by choosing various actions, while assuming that the rest of the extensive-form will be played by backward induction for $(\succ_1, \succ_2)$. A similar construction defines $\mathcal{O}(\nu, \succ_1, \succ_2')$. We prove by backward induction that $\mathcal{O}(\nu_1, \succ_1, \succ_2) \cap \{a, b\} \neq \emptyset$ if and only if $\mathcal{O}(\nu_1, \succ_1, \succ_2') \cap \{a, b\} \neq \emptyset$, for each decision node ν_1 at which the first party makes a choice, and $\mathcal{O}(\nu_2, \succ_1, \succ_2) \cap \{c, d\} \neq \emptyset$ if and only if $\mathcal{O}(\nu_2, \succ_1, \succ_2') \cap \{c, d\} \neq \emptyset$, for each decision node ν_2 at which the second party makes a choice. This is trivially true if these are the last decision nodes. Consider then a decision node ν_1 where the first party makes a decision, and suppose that the property holds true at every subsequent node. We may assume without loss of generality that all the nodes that come right after a decision from the first party are nodes where the second party makes a decision. The second party's optimal action at those nodes leads to an element of $\{c, d\}$ if there is an action that leads to one of these two outcomes when the rest of the subgame is played by backward induction for either $(\succ_1, \succ_2)$ or $(\succ_1, \succ_2')$ (which ever pair of preferences is used to express this condition is irrelevant, thanks to the induction hypothesis). The optimal action at the other nodes does not change when moving from $\succ_2$ to $\succ_2'$ and vice versa, since $\{c, d\}$ is inaccessible for $(\succ_1, \succ_2)$ if and only if it is inaccessible for $(\succ_1, \succ_2')$ (by the induction hypothesis). Hence $\mathcal{O}(\nu_1, \succ_1, \succ_2) \cap \{a, b\} \neq \emptyset$ if and only if $\mathcal{O}(\nu_1, \succ_1, \succ_2') \cap \{a, b\} \neq \emptyset$, as desired. A similar argument applies for a decision node ν_2 at which the second party makes a decision. Take now a node that is reached when the equilibrium strategies for $(\succ_1, \succ_2)$ are followed. The first party has an action that leads to a if the equilibrium path is followed thereafter. This is the best possible option for him, so he has no incentive to take any alternative action if the equilibrium path is followed thereafter, and this is independent of what happens in the subgames that would be reached if he were to choose a different action. The second party also has an action that leads to a if the equilibrium path is followed thereafter. If there was an action that would lead to either c or d when the rest of the game is played thereafter according to the backward induction strategies for $(\succ_1, \succ_2')$, then there would be one that would also lead to c or d when backward induction is applied to $(\succ_1, \succ_2)$ instead, thanks to the property we just proved. No such action exist for the second party along the equilibrium path for $(\succ_1, \succ_2)$, and hence it must be that the second party's action remains optimal when his preference is $\succ_2'$ instead of $\succ_2$, while taking into account that the rest of the game will be played by backward induction according to $(\succ_1, \succ_2')$. Arbitrator a is thus the backward induction outcome of the extensive-form mechanism for $(\succ_1, \succ_2')$. Let $\succ_1'$ be the ordering that coincides with $\succ_1$, except that the ordering of a and b are reversed. The backward induction equilibrium computed for this pair of preferences $(\succ_1', \succ_2')$ must be c, since this is the only element in $f_\alpha(\succ_1', \succ_2')$. A similar reasoning to the one used to move from $(\succ_1, \succ_2)$ to $(\succ_1, \succ_2')$ will imply that c must also be a backward induction outcome of the extensive-form mechanism for $(\succ_1, \succ_2')$. This leads to a contradiction since extensive-form games have a unique backward induction outcome given strict preferences. ■

One alternative SCR is induced by a common sequential mechanism in which the appointed arbitrator is chosen via *alternate strikes* (AS), whereby the two parties take turns removing an arbitrator from the set A until the last remaining arbitrator is chosen. The induced game has a unique SPE outcome, for any pair of preferences. Using a fair coin to decide who is the first mover thus defines a RSCF, ψ_{AS}, and its support defines a SCR, $f_{AS} := support(\psi_{AS})$. Anbarci (2006) provides a characterization of f_{AS} (or equivalently ψ_{AS}).

Proposition 4 *(Anbarci (2006)) $f_{AS}(\succ_1, \succ_2)$ can be computed inductively as follows. Let $A_0(\succ) = A$. For any $B \subseteq A$, let $w_i(B, \succ)$ be player i's least preferred arbitrator. For any integer $t \geq 1$, let*

$$A_t(\succ) = A_{t-1}(\succ) \backslash \{w_1(A_{t-1}(\succ), \succ), w_2(A_{t-1}(\succ), \succ)\}.$$

If t^* is the smallest integer such that A_{t^*} is empty, then $f_{AS}(\succ) = A_{t^*-1}(\succ)$.

It is straightforward to verify that f_{AS} and ψ_{AS} are both Pareto efficient and pass the minimal satisfaction test. Anbarci (2006) also study a natural variant of the alternate strike procedure, that as far as we know is not used in practice. His motivation for introducing such a variant was to derive a SCR that is immune to changes when removing a Pareto inferior arbitrator from the list. One party, call it player 1, proposes an option $a \in A$ to the other party, call it player 2, who may either accept or reject. If 2 accepts, a is chosen; otherwise, a is removed, and player 2 proposes to 1 an option $b \in A \backslash \{a\}$, which player 1 may either accept or reject. The game continues until one of the options is accepted or until only one option remains, which is then chosen. This procedure was studied by Anbarci (2006) under the name of "Voting by Alternating Offers and Vetoes" (VAOV). The induced game has a unique SPE outcome, for any pair of preferences. Using a fair coin to decide who is the first mover thus defines a RSCF, ψ_{VAOV}, and its support defines a SCR, $f_{VAOV} := support(\psi_{VAOV})$.

Proposition 5 *(Anbarci (2006))* $f_{VAOV}(\succ_1, \succ_2) = \arg\max_{a \in A} \min_{i=1,2} \sigma(\succ_i, a)$*, where* $\sigma(\succ_i, a) = \#\{a' \in A | a \succ_1 a'\}$.

It turns out that this SCR has already been studied previously in the literature under a variety of names: "Rawlsian arbitration rule" (Sprumont (1993)), "Kant-Rawls Social Compromise" (Hurwicz and Sertel (1997)), "fall-back bargaining" (Brams and Kilgour (2001)), as well as "unanimity compromise" (Kibris and Sertel (2007)). We adopt the name suggested by Sprumont (1993). It is easy to check that f_{VAOV} and ψ_{VAOV} are both Pareto efficient and pass the minimal satisfaction test. Notice that the SCRs f_{VAOV} and f_α share the common feature of using scores based on the two parties ordinal rankings, a tradition that goes back at least to the 18^{th} century with Borda. While f_α uses these scores in a utilitarian tradition, summing them up, f_{VAOV} uses them in an egalitarian tradition, aiming at maximizing the welfare index of the worse-off party. Vetoes were needed for the SCR f_α to pass the minimal satisfaction test (when $\alpha = (n-1)/2$). Applying the egalitarian criterion instead guarantees that the resulting SCR passes that test without the need to resort to vetoes.

It is well documented that backward induction does not systematically prevails when the game has multiple stages (see e.g. Binmore et al. (2002) and Levitt, List and Sadoff (2010)). There are thus reasons to focus on short extensive forms. We now investigate the simplest possible such mechanisms. Formally, a *two-stage mechanism* is composed of a finite set of actions A_1 for the first mover, the identity of which can be chosen by tossing a fair coin, a function A_2 that determines a finite set of actions for the second mover (the other party), and an outcome function o that selects an element in A for each pair (a_1, a_2) such that $a_1 \in A_1$ and $a_2 \in A_2(a_1)$. Backward induction leads to an optimal strategy for the second mover: $a_2^*(a_1, \succ_2) \in A_2(a_1)$, for each $a_1 \in A_1$ and each $\succ_2 \in \mathcal{P}$, such that $o(a_1, a_2^*(a_1, \succ_2))$ is optimal according to $\succ_2$ within $\{o(a_1, a_2) | a_2 \in A_2(a_1)\}$. Then the optimal strategy for the first mover is given by $a_1^*(\succ)$ where $o(a_1^*(\succ), a_2^*(a_1^*(\succ), \succ_2))$ is optimal according to $\succ_1$ within $\{o(a_1, a_2^*(a_1, \succ_2)) | a_1 \in A_1\}$, for each $\succ \in \mathcal{P} \times \mathcal{P}$. Let $o^* : \mathcal{P} \times \mathcal{P} \to A$ be the outcome of the two-stage mechanism when played by backward induction:

$o^*(\succ) = o(a_1^*(\succ), a_2^*(a_1^*(\succ), \succ_2))$, for each $\succ \in \mathcal{P} \times \mathcal{P}$. The function o^* is the SCR implemented by the two-stage mechanism (A_1, A_2, o).

Proposition 6 *There exists a unique single-valued SCR o^* that is Pareto efficient, passes the minimal satisfaction test, and can be implemented by backward induction via a two-stage mechanism. It is computed as follows:*[7]

$$o^*(\succ) = \arg\max_{\succ_1}\{a \in A | \#\{b \in A | a \succ_2 b\} \geq \lfloor \frac{n-1}{2} \rfloor\}.$$

In addition, the following two-stage "shortlisting" mechanism implements o^*: A_1 is the set of subsets of A with $\lfloor \frac{n+1}{2} \rfloor$ elements, $A_2(a_1) = a_1$, for each $a_1 \in A_1$, and $o(a_1, a_2) = a_2$.

Proof: It is easy to check that the two-stage shortlisting mechanism proposed implements o^*, and that o^* is Pareto efficient and passes the minimal satisfaction test. Hence we will limit ourselves to prove that o^* is the only SCR with those properties. Let $\succ \in \mathcal{P} \times \mathcal{P}$. We now define a new ordering $\succ_1'$ for the first mover. First the elements ranked above $o^*(\succ)$ according to $\succ_1$ keep the same rank[8] in $\succ_1'$. Notice that the rank of all these elements must be strictly larger than $\lfloor \frac{n+1}{2} \rfloor$ in $\succ_2$, by definition of o^*. Then place the other elements ranked strictly larger than $\lfloor \frac{n+1}{2} \rfloor$ in $\succ_2$ (if any) in some specific order (let's say alphabetically) in the next available spots in $\succ_1'$ (that is, after those elements above $o^*(\theta)$ according to $\succ_1$). The next available spot in $\succ_1'$ must be the $\lfloor \frac{n+1}{2} \rfloor$-rank. Place $o^*(\theta)$ there, and then rank the remaining elements in some specific order (let's say alphabetically again). Let $\bar{o}$ be a single-valued SCR that can be implemented via a two-stage mechanism, is Pareto efficient and passes the minimal satisfaction criterion. The minimal satisfaction test applied to both players implies that $\bar{o}(\succ_1', \succ_2) = o^*(\succ)$. Notice that the lower contour set of $o^*(\succ)$ expands when moving from $\succ_1'$ to $\succ_1$. Hence the backward induction outcome of the two-stage mechanism in $(\succ_1, \succ_2)$ must be the same

[7]For each positive real number x, $\lfloor x \rfloor$ will denote the largest integer that is no larger than x.

[8]The rank of a top ranked element is 1. The rank of the second element according to the ordering is 2, and so on so forth. The rank is thus equal to n minus the score.

as the one in $(\succ_1', \succ_2)$ (the second party's optimal strategy remains unchanged since his preference remains fixed), or $\bar{o}(\succ) = \bar{o}(\succ_1', \succ_2)$. We get $\bar{o}(\succ) = o^*(\succ)$ by transitivity, as desired. ■

Considering the two procedures derived with the first or the second party is the first mover leads a natural SCR that is fully implementable, and picking which assumes the role of the first mover by using a fair coin leads to an RSCF that is implementable.

While truthtelling is not always an equilibrium strategy for the first mover (i.e., selecting his top $(n+1)/2$ elements), an equilibrium strategy (or a best response to the belief that his opponent is rational) can be derived using the following simple algorithm. In the first step, player 1 checks if there is a set S of $(n-1)/2$ elements that player 2 ranks below 1's top choice, a. If so, 1 chooses $\{a\} \cup S$. Otherwise, he goes to the next step. In the second, step player 1 checks if there is a set T of $(n-1)/2$ elements that 2 ranks below 1's second-best choice, b. If so, 1 chooses $\{b\} \cup T$. Player 1 continues in this fashion until the algorithm terminates at or before 1's median choice.

4 Empirical Analysis

As a complement to the experimental analysis developed in the next two sections, we start by presenting results from an empirical analysis based upon real-world data in which employers and unions had to select an arbitrator. Arbitration cases typically involve both wages and benefits, and the returns from selecting a favorable or high-quality arbitrator to the union and the employer can thus be quite high. Thus, one advantage of the empirical analysis involves the high stakes in the selection of an arbitrator.

In particular, we use information from the New Jersey Public Employment Relations Commission (PERC). During the years 1985 to 1996, employers and unions were provided a menu of seven arbitrators and were asked to veto three arbitrators and to rank the remaining four.[9] The arbitrator with the lowest

[9]After 1996, the an arbitrator is randomly selected by a computer from the list of approved arbitrators.

combined rank among those that were not vetoed by either party was then chosen as the arbitrator for the case. This mechanism thus corresponds to the normal-form mechanism with $n = 7$ and $v = 3$.

4.1 Data

Data on rankings by employers and unions were provided by the New Jersey Public Employment Relations Commission (PERC) and cover the years 1985 to 1996. Variables in this dataset include the year of the case, the names of the two parties (employer and union), the menu of arbitrators (including first and last name), the rankings of each party, and the name of the arbitrator chosen by the procedure. Employers in these data are local governments within the state of New Jersey. These include municipalities, such as the city of Trenton, agencies within municipal governments, such as corrections in Middlesex County, and agencies within the state government, such as the New Jersey State Police. Unions represented are then public sector unions within the relevant government or government agency.

We drop a number of cases with inconsistent or incomplete data. Examples of cases with inconsistent or incomplete data include some cases in which the arbitrator chosen does not reflect the mechanism described above, cases in which one or both of the two parties did not submit a ranking, and cases in which parties submitted rankings but did not follow the request to veto three options and rank the remaining four. After deleting these observations, we are left with 750 cases with complete rankings by employers and employees and in which the chosen arbitrator followed the rankings submitted by the two parties. Given that the menu includes seven arbitrators, we thus have 5,250 arbitrator choices and 10,500 unique ranks, one for the employer and one for the union.

The data are a panel in the sense that there is overlap across cases in terms of arbitrators, employers and unions. While there are 5,250 arbitrator choices across the 750 cases, we find only 101 unique arbitrators represented, with one arbitrator appearing in 256, or over one-third, of the cases. We also utilize the fact that employers are frequently represented in multiple cases. In particular, there are 350 unique employers represented in the data, and a typical employer will thus be represented in two or three different cases. At the extreme, one employer is represented in fifteen different cases. Unions are also repeated in the data, although to a lesser extent than employers. In particular, there are 419 unique unions represented in the data and hence fewer cases of repeated observations.

We unfortunately have limited information on the characteristics and preferences of the employer, union, and arbitrators. For example, due to the substantial amount of time elapsed since the conclusion of these cases, we were unable to locate the resumes of the arbitrators that were analyzed in Bloom and Cavanagh (1986a). Given this, we cannot infer the preferences of the two parties over the characteristics of arbitrators and thus cannot infer directly whether or not the strategies that were played reflect the preferences of the relevant parties. As will be seen below, we thus rely on a test of pairwise reversals in rankings among employers.

4.2 Tests for Perfect Opposition of Preferences

As noted in the introduction, an important issue involves whether preferences are completely opposed. To shed light on this issue, we examine the rankings and vetoes of both the employer and the union. When preferences are perfectly opposed, truthful behavior is a Nash equilibrium. Moreover, truthful play may be considered the focal equilibrium in this case. Thus, if preferences were perfectly opposed, then there should be no overlap in terms of either the rankings or the vetoes.

As shown in Table 1a, however, we see little evidence of perfect opposition in terms of rankings.[10] For example, conditional on an arbitrator being ranked first by the union, this arbitrator is also ranked first by the employer in 17 percent of cases, ranked second in 15 percent of cases, ranked third in 15 percent of cases, ranked fourth in 11 percent of cases, and vetoed in 42

[10]We can also use this evidence to argue that preferences are not perfectly aligned.

		Union ranking					
	Profile	1	2	3	4	veto	total
Employer ranking	1	125	96	115	89	325	750
		2.38%	1.83%	2.19%	1.70%	6.19%	14.29%
	2	112	123	112	96	307	750
		2.13%	2.34%	2.13%	1.83%	5.85%	14.29%
	3	112	105	99	102	332	750
		2.13%	2.00%	1.89%	1.94%	6.32%	14.29%
	4	85	111	100	118	336	750
		1.62%	2.11%	1.90%	2.25%	6.40%	14.29%
	veto	316	315	324	345	950	2,250
		6.02%	6.00%	6.17%	6.57%	18.10%	42.86%
	total	750	750	750	750	2,250	5,250
		14.29%	14.29%	14.29%	14.29%	42.86%	100.00%

Table 1: Distribution of union and employer rankings of arbitrators

percent of cases. The nearly uniform distribution across these categories is thus inconsistent with perfect opposition.

One limitation of this test involves multiplicity of equilibria. In particular, when preferences are perfectly opposed, parties always veto their bottom $(n-1)/2$ ranked options in equilibrium but any ranking of the remaining $(n+1)/2$ options constitutes a Nash equilibrium. Indeed, we will see in our experimental analysis to follow that subject pairs are truthful in only 25% of cases when preferences are completely opposed, and this is due mainly to deviations from sincerity on the ranking of the non-vetoed options.

To address this issue, we next develop a more robust test based upon the fact that, as noted above, there should be no overlap in vetoes when preferences are perfectly opposed. Based upon an analysis of vetoes, however, we find a substantial degree of overlap in vetoes. In particular, in 50% of cases there is one common veto, in 34% of cases there are two common vetoes, and in 3% of cases there are three common vetoes. That is, in 87% of the cases the parties' vetoes overlap, and there is no overlap in only 13% of cases. Taken together, these tests find little evidence that preferences are perfectly opposed.

While this descriptive analysis is useful in terms of evaluating this extreme case of perfect opposition, it does not shed light directly on the question of whether players are strategic or sincere. To test for strategic play, we next turn to an analysis of preference reversals.

4.3 Test for Strategic Behavior

Our test of strategic play exploits the panel aspect of the data and involves an analysis of pairwise reversals in rankings. If an employer, for example, ranks arbitrator a over b in one case but arbitrator b over a in another case, we infer that, under an assumption of stable preferences, these rankings do not reflect the preferences of the employer and hence that the employer was strategic.

As should be clear, this test requires an assumption of complete information, as maintained above, and an additional assumption of stable employer preferences for a given employer across cases. Under this assumption, if the same pair of arbitrators appears in two different cases, then one should always be ranked above the other by non-strategic players with the same preferences. Pairwise reversals in rankings would thus provide evidence of strategic behavior.

This test is aided by the fact that, as noted above, the same arbitrators appear repeatedly in the data. Given that employers occur in multiple cases more commonly than do unions, we focus our analysis on comparing employer rankings of arbitrator pairs across multiple cases. Based upon these repeated occurrences of arbitrators and employers, we found 447 observations in which an employer had the same two arbitrators in the choice set in two different arbitration cases and in which the two sets of rankings can be compared.[11]

One reason that the assumption of stable preferences might be violated involves learning by employers about the desirability of different arbitrators. We address this issue by excluding observations in which one of the arbitrators was assigned to the employer in the period between the two cases. In these situations, it is reasonable to assume that the employer learns something, such as the tendency of this arbitrator to side with the employer, from this experience. Such learning could potentially change the employer's preferences over arbitrators in the context of future cases. Of the original sample of 447 observations, we find that one of the arbitrators was assigned to the employer

[11] If both arbitrators are vetoed in one (or both) of the two cases, we cannot determine the relative ranking of these two options in that case, and these observations thus excluded from the analysis.

	consistent	reversal	total
no experience	165	84	249
	66.27%	33.73%	
experience	111	87	198
	56.06%	43.94%	

Table 2: Pairwise reversal of employer rankings by experience with arbitrator

in 198 observations in the period between the two cases.[12] Excluding these 198 observations, we have 249 observations in which the employer had no interaction with the two arbitrators during the period between the two cases. In these cases, we do not expect employers to learn something about the arbitrator, and we thus assume that employer preferences over arbitrators are stable in these cases.

As shown in Table 2, of these 249 observations, the relative rankings of the two arbitrators switches in around one-third (34 percent) of the observations. The relative rankings of the two arbitrators is unchanged/consistent in the other two-thirds (66 percent) of the observations. This finding of a substantial number of reversals is inconsistent with sincere rankings under the assumption of stable preferences and suggests that there is some element of strategic play involved in these rankings.

As noted above, this analysis excludes the observations in which the employer had some experience with the arbitrator during the period between the two cases. For comparison purposes, we next present results for this set of observations. As shown in Table 2, among the 198 observations with interactions during the time period between the two cases, we see a somewhat higher switching rate (44 percent versus 34 percent), and this difference of 10 percentage points between these two types of cases is statistically significant at the 95-percent level. This finding of a higher switching rate when the employer had an interaction with the arbitrator is consistent with employers learning from working with specific arbitrators.

As noted above, this test of pairwise reversals requires an assumption of

[12] This could be due to the arbitrator being selected in the first of the two cases or in a separate case during the intervening time period.

		consistent	reversal	total
time between cases	0	44	9	53
		83.02%	16.98%	1.01%
	1	25	12	37
		67.57%	32.43%	0.70%
	2	29	16	45
		64.44%	35.56%	0.86%
	3	15	11	26
		57.69%	42.31%	0.50%
	4	20	13	33
		60.61%	39.39%	0.63%
	5	5	3	8
		62.50%	37.50%	0.15%
	6	8	5	13
		61.54%	38.46%	0.25%
	7	9	6	15
		60.00%	40.00%	0.29%

Table 3: Pairwise reversal of employer rankings by time between cases

stable preferences in the two cases. To shed light on the validity of this assumption, we next use information on the time that elapsed between the two cases. It is reasonable to assume that if preferences are not stable, then more switching in rankings should occur as the time that elapsed between the two cases increases.

As shown in Table 3, which focuses on the subset of cases in which the employer had no interaction with either of the two arbitrators under consideration in the period between the two cases, switching rates are increasing as the number of years between the two cases increases. In particular, if the cases occur during the same year, then reversals occur in around 17 percent of cases. This rate increases to 32 percent if there is one year between the two cases and to 36 percent if there are two years between the two cases. Beyond two years, the reversal rates stabilize at around 40 percent.[13]

This finding of an increase in reversal rates as the elapsed time increases suggests that the assumption of stable preferences is questionable. On the other hand, there are still a sizeable fraction of switches, 17 percent, when the two cases occur during the same year. It seems unlikely that preferences would

[13]While there could be up to 11 years between the two cases, the number of observations is too small to conduct a meaningful analysis when the number of years between the two cases exceeds seven.

change during the same year, especially given that the employer has no interaction with the arbitrators in these cases. Thus, while we find some evidence that preferences are not stable, we continue to see instances of switching in cases where the assumption of stability is most plausible. Thus, any possible instability in preferences cannot entirely explain the finding of a substantial number of switches in Table 2.

While this empirical analysis has the advantage of a high-stakes environment with rankings by professionals, we cannot evaluate the key assumptions of complete information and stable preferences. Given this, we next turn to an experimental analysis. While the stakes are lower in this environment and players are students, we can control payoffs and information in this experimental setting. Given this, we view these two approaches as complementary.

5 Experimental design

According to the theory, the prevalent normal-form mechanism for choosing arbitrators is plagued with multiple equilibria, some of which lead to "undesirable" outcomes in the sense that they do not maximize the sum of the parties' (canonical) utilities ("scores") over the set of arbitrators, which are better than at least k of the arbitrators. In addition, for most preference profiles, truthful ranking is not a Nash equilibrium. The multiplicity of equilibria also raises the problem of miscoordination: the participants in this mechanism may each choose a strategy belonging to a different Nash equilibrium. Such miscoordination may result in the selection of Pareto inferior arbitrators. For example, when the preference profile is given by $a \succ_1 b \succ_1 c \succ_1 d \succ_1 e$ and $b \succ_2 a \succ_2 c \succ_2 d \succ_2 e$, there is a Nash equilibrium in which player 1 vetoes b and another equilibrium in which player 2 vetoes a. But if the two players choose these strategies, the resulting arbitrator would be Pareto dominated by both a and b.

Of course, none of these issues would be of any concern if the two parties who actually participate in this mechanism do not behave strategically as the theory assumes. In particular, the mechanism would attain desirable outcomes if the parties naïvely delete the bottom k options and truthfully report their ranking for the remaining arbitrators.

In contrast, the sequential mechanisms we discussed, induce equilibrium outcomes, which are all desirable, in that they implement closely related SCRs/RSCFs and satisfy the basic properties of Pareto efficiency and minimal satisfaction. This result relies on the participants' abilities to perform backward induction. However, a number of studies in the experimental literature suggest that most subjects find it difficult to perform backward induction, and often fail to carry it out.

Thus, in order to evaluate the performance of the mechanisms described in Section 3, it is important to understand how the participants in these mechanisms would actually behave. We, therefore, conducted a series of computerized laboratory experiments that test these mechanisms. The experiments were conducted at NYU's Center for Experimental Social Science. A total of 304 subjects from the undergraduate student population participated.

In each treatment, an even number of subjects was presented with a set of five alternatives, $A = \{a, b, c, d, e\}$, and were randomly matched to play one of the mechanisms on this set of options. Each treatment consisted of 40 rounds, which were divided into four "blocks" of ten rounds. In each of these blocks, subjects had the same preference relation over the five options, but these preferences changed from one block to another (i.e., in total there are four distinct preference profiles). Preferences over A are induced by assigning each of the options a distinct monetary value in the set $\{\$1.00, \$0.75, \$0.50, \$0.25, \$0.00\}$. The four preference profiles were as follows

1	2
a	e
b	d
c	c
d	b
e	a

p_1

1	2
a	b
b	a
c	c
d	d
e	e

p_2

1	2
a	c
b	b
c	a
d	d
e	e

p_3

1	2
a	e
b	c
c	a
d	b
e	d

p_4

The first profile (denoted p_1) consists of completely opposed rankings; the second profile, p_2, represents partial conflict of interest involving only the top two options; the third profile, p_3, represents partial conflict of interest at the top with a focal compromise; and the fourth profile, p_4, which we discussed above.

There were four treatments, each corresponding to one of the mechanisms we discussed. There were 70 participants in the first treatment, 74 in the second, 72 in the third and 88 in the fourth. For each mechanism and each preference profile, we have characterized the set of pure-strategy equilibria.[14] For each treatment we ran four sessions, where in each session the four induced preference profiles appear in a different order. The four orders were: $p_1p_2p_3p_4$, $p_4p_3p_2p_1$, $p_1p_3p_2p_4$ and $p_4p_2p_3p_1$. Hence, each profile was played (by a different group of subjects) at two different stages in the experiment: an "early" stage (the first ten rounds for p_1 and p_4 and the second block of ten rounds for p_2 and p_3) and a "late" stage (the last ten rounds for p_1 and p_4 and the third block of ten rounds for p_2 and p_3). This allows us to examine whether there was a learning "spillover" from one profile to another.

Subjects were paid the sum of their earnings across the 40 rounds in addition to a show-up fee of \$10. Appendix 1 contains the instructions to one of the treatments (instructions to the other treatments were similar and are available from the authors upon request). After subjects read the instructions they were presented with a short quiz that tested their understanding of the game. When the subjects finished answering the quiz, they were presented with the correct answers. The instructions in the appendix also include the quiz that followed them.

[14] It is straightforward to verify whether a pair of actions constitute an equilibrium in the rank-veto and the shortlisting games. The characterization for AS and VOAC is more involved and is available from the authors upon request. Equilibrium strategies for the shortlisting scheme is described in Section 3.

6 Experimental results

In this section we report on key observations from the data of each of the treatments. Tests of significance that are reported in this section take into account correlation across rounds.

6.1 Simultaneous move game

6.1.1 Outcomes

As discussed above, the simultaneous-mechanism is appealing as long as the participants are truthful in the sense that a participant vetoes his bottom two options and ranks the remaining alternatives according to his true preferences. To shed light on this issue, Table 4 displays, for each preference profile, the frequency of outcomes. The light shaded cells represent the outcomes with the highest observed frequency, the shaded cells represent the outcomes corresponding to truthfulness, and cells with a dark outline represent outcomes that have the highest frequency and that also correspond to truthfulness. When two letters are written adjacent to each other (e.g., ab), we mean a tie between these outcomes.

The results suggest that a large fraction of players are *not* truthful. While sincere play should result in a tie between a and b in profile p_2, the total frequency of either a or b being chosen (both of which are Nash equilibrium outcomes) is greater than the mean proportion of the truthful prediction, ab, and this difference is statistically significant at the 1% level. Similarly, in p_3, there should be a greater mass on ties (between any subset of elements in $\{a, b, c\}$) under sincere play than on any individual outcomes. The results demonstrate, by contrast, that the mean proportion of b, a Nash equilibrium outcome, is significantly higher than that of the truthful prediction abc (significant at the 1% level). Finally, while sincere play results in a being chosen with certainty under p_4, we find a large fraction, roughly one-quarter, of cases resulting in c, which is a Nash equilibrium outcome.

outcome	mean	p1	p2	p3	p4
a	early	0.07	0.24	0.08	0.37
	late	0.03	0.19	0.16	0.41
	pooled	**0.05**	**0.21**	**0.12**	**0.39**
b	early	0.09	0.47	0.38	0.08
	late	0.05	0.43	0.38	0.12
	pooled	**0.07**	**0.45**	**0.38**	**0.10**
c	early	0.54	0.02	0.11	0.30
	late	0.87	0.05	0.11	0.18
	pooled	**0.70**	**0.03**	**0.11**	**0.24**
d	early	0.08	0.01	0.06	0.07
	late	0.01	0.03	0.08	0.06
	pooled	**0.05**	**0.02**	**0.07**	**0.06**
e	early	0.06	0.01	0.03	0.01
	late	0.01	0.00	0.04	0.02
	pooled	**0.03**	**0.00**	**0.03**	**0.01**
ab	early	0.02	0.20	0.10	0.01
	late	0.00	0.23	0.07	0.02
	pooled	**0.01**	**0.21**	**0.08**	**0.01**
abc	early	0.00	0.03	0.11	0.01
	late	0.01	0.02	0.11	0.01
	pooled	**0.00**	**0.03**	**0.11**	**0.01**
ac,bc	early	0.03	0.02	0.11	0.14
	late	0.02	0.05	0.03	0.15
	pooled	**0.02**	**0.03**	**0.07**	**0.15**
other	early	0.10	0.01	0.04	0.02
	late	0.02	0.01	0.02	0.04
	pooled	**0.06**	**0.01**	**0.03**	**0.03**

Table 4: Mean proportions of outcomes in simultaneous mechanism

6.1.2 Strategies

To better understand our subjects' behavior we turn next to analyze their actions. We first consider only the veto choices of our subjects. The veto matrices for each preference profile are displayed in Appendix 2. These matrices give us some direct evidence on the extent to which subjects were strategic: in profiles where the two players agree on the dominated options (as in p_2 and p_3), each player has an incentive to remove his opponent's top option (relying on the opponent to remove the dominated outcomes). In contrast, truthtelling calls for both players to remove their two least preferred options.

In profile p_2, our data reveal that a majority of subjects did "veto truthfully" (i.e., they removed their least preferred options). However, around one-third of subjects in the role of player 2 removed their opponent's top option, a, and about 15% of subjects in the role of player 1 removed player 2's top pick, b.[15]

The veto choices of the subjects can also be described within the framework of k-level reasoning (see the survey in Crawford et al. 2010). A natural candidate for level zero behavior is being truthful. Level 1 would then constitute the best response against truthful behavior. In about 26% of the matched pairs, player 1 behaved according to level zero and vetoed $\{d, e\}$, while player 2 behaved as level 1 and vetoed a. Only 8% of the pairs exhibited the reverse behavior: player 2 being level zero and player 1 being level 1. Only about 5% of matched subjects behaved as level 1, thereby inducing an inefficient outcome.

In p_3, the majority of subjects also vetoed truthfully. Around 40 percent of subjects, by contrast, strategized by removing the opponent's top option. In addition, almost one-quarter of pairs contained a subject who behaved as level zero (vetoes bottom two options) and a subject who behaved as level one (vetoed opponent's top pick). The proportions for player 1 and player 2 were similar. Finally, about 11% of the pairs selected an inefficient outcome because each player removed his opponent's top element, and at least one also removed the opponent's second best option.[16]

When faced with p_4, truthtelling implies that player 1 vetoes $\{d, e\}$ and player 2 vetoes $\{b, d\}$, and around one-half of subjects followed these strategies. Player 1 is strategic if he vetoes outcome c, which is the second-best for his opponent and third-best for himself, and we find that 42% of subjects in the role of player 1 follow this strategy. The proportion of subjects who behaved strategically as player 2 (veto a, which is best for player 1 and third-best for 2) is 37%.[17] Interestingly, player 2 vetoes his most preferred alternative (e) in around 7% of cases, and, not shown here, ranks this option first in only 28% of cases, providing evidence of a violation of the key assumption in Bloom and

[15] The source of this asymmetry between the two players is not clear to us. One hypothesis is it was caused by framing: player 1's ranking of the options coincided with the natural ordering of the alphabet, which may have served as a salient focal point.

[16] This percentage is significantly different form zero at the 1% level.

[17] This does not include less than 1% of subjects who as player 2 removed the top elements of *both* players.

Cavanagh (1986a). The proportions of pairs in which one player was level zero and another was level 1 were 25% when player 1 was level 1 and 21% when player 2 was level 1. Miscoordination occurs between the players in about 15% of cases when both act as level 1 such that outcomes a, c and e are removed and only dominated outcomes remain.

Since players can also strategize through the ranking (e.g., instead of removing the opponent's top option - even when that option is not dominated - lie about its ranking), we turn next to consider the entire action chosen by each subject. We first categorize action profiles into seven broad categories: (*i*) *truthful play* (or level zero) - each player removes his worst two options and truthfully ranks the remaining items, (*ii*) *Nash* - a pair of actions that constitute a Nash equilibrium, (*iii*) *level 01* - one player being truthful and the other player best responding, (*iv*) *level 11* - each player best responding to a belief that the opponent is truthful (*v*) level 12 - one player's action is coincides with level 1 who best responds to a truthful player and another player is level 2 in that he best responds to a level 1 player, (*vi*) *level 22* - each player best responds to a belief that his opponent is level 1, and (*vii*) *other* - play paths that did not belong to any of the previous categories. The categories are organized such that they are mutually exclusive: any action pair which coincides with truthful play is labeled truthful - *even if it is consistent with another category*, and similarly, any non-truthful action pair, which constitutes a Nash equilibrium is classified as Nash - *even if it is consistent with another category*. If a certain action fits more than one level in the k-level hierarchy, we label it with the smallest level.

Table 5 displays the mean proportion of each category for each of the preference profiles and provides further evidence for strategic behavior.

As shown, for every preference profile, and for every block of ten rounds, less than a third of matched pairs are truthful, and less than 8% of pairs fall into the "other" category. This means that the behavior of at least 60% of pairs can be explained as resulting from strategic behavior: either a Nash equilibrium or a combination of level-k behavior. In p_1 it is a Nash equilibrium for each player to be truthful. Hence, in a sense, p_1 is the profile where we

Profile	Mean	truthful	nash	level10	level11	level12	level22	other
	1-10	0.19	0.32	0.00	0.01	0.33	0.07	0.08
1	31-40	0.32	0.49	0.04	0.01	0.14	0.00	0.00
	Pooled	**0.25**	**0.40**	**0.02**	**0.01**	**0.24**	**0.04**	**0.04**
	11-20	0.16	0.36	0.17	0.08	0.05	0.18	0.00
2	21-30	0.16	0.33	0.13	0.11	0.08	0.18	0.01
	Pooled	**0.16**	**0.35**	**0.15**	**0.09**	**0.06**	**0.18**	**0.01**
	11-20	0.10	0.12	0.03	0.03	0.14	0.59	0.00
3	21-30	0.10	0.24	0.04	0.04	0.10	0.47	0.00
	Pooled	**0.10**	**0.19**	**0.04**	**0.03**	**0.12**	**0.53**	**0.00**
	1-10	0.06	0.35	0.04	0.12	0.22	0.21	0.00
4	31-40	0.05	0.39	0.02	0.11	0.23	0.17	0.03
	Pooled	**0.05**	**0.37**	**0.03**	**0.11**	**0.23**	**0.19**	**0.01**

Table 5: Mean proportions of action pairs in simultaneous mechanism

would expect the most truthful behavior. Yet, only 25% of pairs were truthful (mean proportion over all 20 rounds), while 40% of pairs coordinated on a non-truthful equilibrium. In p_2 truthtelling is not a Nash equilibrium, and the proportion of Nash-playing pairs is more than twice that of truthful pairs. Similarly, in p_3 there are twice as many pairs playing non-truthful Nash than there are pairs being truthful. Interestingly, more than half of the pairs could be described as being players of level 2. Finally, in p_4 about 37% of the pairs play a non-truthful Nash equilibrium, while only 5% of the pairs are truthful.[18]

6.2 Alternate strikes

6.2.1 Outcomes

Table 6 displays the distribution of outcomes for each of the preference profiles. For p_1 it is clear that the vast majority of times (84%) the selected outcome was c, the unique SPE outcome.[19] The unique SPE outcome for p_2 is the first

[18]Table 5 allows us to test whether subjects learned across preference profiles. In p_1 the proportion of both truthful behavior and (non-truthful) Nash behavior is higher in rounds 31-40 than in rounds 1-10 (significant at the 1% level). Non-truthful Nash behavior also increases when p_3 is played in rounds 21-30 compared when it is played in rounds 11-20. However, there is no change in the frequency of truthful behavior. In the remaining cases both the frequency of truthful and (non-truthful) Nash behavior remains unchanged when a profile is played at a later stage in the game.

[19]Looking at the proportion of outcomes across the different rounds the following observations stand out: (1) there is a clear gap between the outcome c and all other outcomes, (2) this gap widens when the profile is played in the last ten rounds. When subjects face p_1

Profile	Mean	a	b	c	d	e
1	1-10	0.02	0.11	0.78	0.08	0.02
	31-40	0.00	0.06	0.91	0.04	0.00
	Pooled	**0.01**	**0.08**	**0.84**	**0.06**	**0.01**
2	11-20	0.44	0.36	0.15	0.04	0.01
	21-30	0.42	0.43	0.14	0.01	0.00
	Pooled	**0.43**	**0.39**	**0.15**	**0.02**	**0.01**
3	11-20	0.09	0.79	0.08	0.04	0.01
	21-30	0.05	0.85	0.06	0.03	0.01
	Pooled	**0.07**	**0.82**	**0.07**	**0.03**	**0.01**
4	1-10	0.51	0.09	0.32	0.03	0.06
	31-40	0.51	0.11	0.33	0.05	0.01
	Pooled	**0.51**	**0.10**	**0.32**	**0.04**	**0.03**

Table 6: Mean proportions of outcomes in AS

player's top element. Since in our experiment each player had an equal chance of being selected to be the first mover, the theoretical prediction is that only a and b should be selected, and in equal frequency. The mean proportions of a and b are 43% and 39%, respectively, and this difference is not statistically significant. The mean proportion of c (which is dominated by both a and b) is about 15%.[20]

The relatively high proportion of c outcomes is surprising since, in contrast to the simultaneous mechanism, miscoordination cannot occur in this sequential game. One possible explanation may be due to social preferences. The mere fact that player i eliminates - on his own initiative - player j's top alternative may induce j to "get even" by removing i's top alternative, even though this is against his own material interests.[21] To explore the above hypothesis, we identified all the instances in which player i removed the top alternative of player j while i's top alternative was still available and calculated the proportion of these instances where j responded by removing i's most preferred

in the last ten rounds (as opposed to the first ten rounds), the frequency of c rises (p level of 1%), while the frequency of all other options declines (p level of 3%)).

[20]There is a clear gap between the equilibrium outcomes and the rest of the outcomes. When p_2 is played in rounds 21-30 as opposed to rounds 11-20, the proportion of the dominated options d and e decreases (5% significance), while the proportion of the remaining outcomes is unchanged.

[21]Framing may also play a role because - as we show later - it appears that the alternating offer game does not induce this type of behavior. That is, when i rejects a proposal to implement i's top alternative it is equivalent (in terms of outcomes) to i eliminating j's top choice. However it may not be perceived in the same way by the players. One reason for this may be that it seems only fair that a greedy proposal by one of the players is not accepted.

remaining alternative. Around 20 percent of players behaved in this way, lending some support to our hypothesis (the percentage drops from about 23% in rounds 11-20 to about 16%, and this decrease is significant at the 10% level).

For p_3 the unique SPE outcome is b regardless of who moves first. The data appears to support this prediction as the mean proportion of b is more than 80% in the aggregate data. There are two dominated alternatives, d and e, and their mean proportion in the aggregate pool (over the four sessions) is less than 3%.[22]

When payoffs are given by p_4 the unique SPE outcome is a, regardless of which player moves first. While a was selected by our subjects in a majority of cases, other outcomes were also selected with non-negligible frequency. In particular option c was selected 32% of the times, and b, which is Pareto dominated by a, was selected 10% of the times.[23] The frequencies of these outcomes appears to be rather stable across the rounds and between the first and last ten rounds.[24]

6.2.2 Strategies

The theoretical analysis of AS assumed that players are able to perform backward induction over any length of histories. To get a sense of whether our subjects could perform backward induction at some "minimal" level, we examined how subjects behaved when they were the penultimate movers, i.e. when there were three options left. If at this stage, a player believes that his opponent is rational, then the player expects the opponent to eliminate his lower ranked alternative of the remaining two.[25] Figure 1 shows the fraction

[22] In addition, there is a clear gap between the sequence of proportions of b across the rounds and the sequence of aggregated proportions of the other outcomes. The frequencies of the outcomes is the same whether p_3 is played in the second or third block of ten rounds.

[23] The difference between a and c is statistically significant at the 1% level, as well as the difference between each of these outcomes and the remaining ones.

[24] The only exception is the frequency of e, which is lower in the last ten rounds (p level of 1%)

[25] More formally, for any triplet of alternatives S, and for any $x \in S$, let $y_j(S \backslash \{x\})$ be player j's worst element in $S \backslash \{x\}$. It follows that if player i is rational, he should choose x to maximize his preference over $S - \{x\} - y_j(S \backslash \{x\})$.

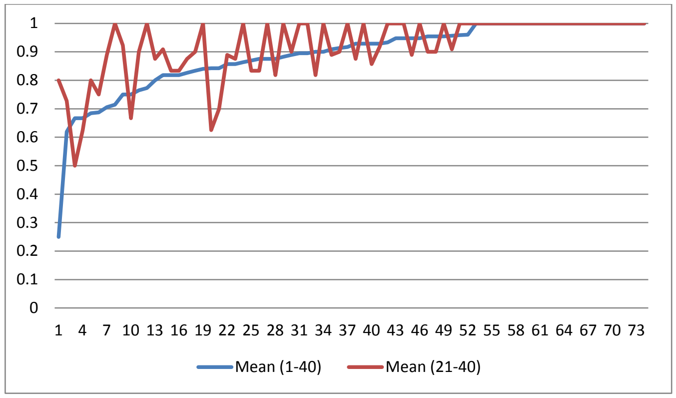

Figure 1: Backwards induction in AS - Mean proportion of times a player's action was consistent with backwards induction when only three items remained.

of times that each penultimate mover behaved according to this prediction. The horizontal axis depicts the subject ID, and the vertical axis depicts the proportion of times a subject was doing backward induction. Subjects have been rearranged so that the curve representing the mean for rounds 1-40 is increasing. Learning is observed whenever the curve representing rounds 21-40 is above that of rounds 1-40. As is clear from the figure, most subjects were able to perform this minimal level or backward induction. In addition, this ability seems to rise as subjects gain experience in playing the mechanism under different preference profiles.

To better understand how subjects played the game, we examined the paths played by the matched pairs. We classified these paths into six categories.

1. *SPE* play - A path induced by some pair of SPE strategies.

2. *Truthful* play (or level 0) - Each player removes his bottom ranked option among those remaining.

3. *Level one and zero* (level 10) - The first mover best responds to a belief that second mover plays truthfully, and second mover indeed plays truthfully.

4. *Level one and spe* (level 1s) - The first mover's initial action coincides with the initial action of a best response to truthful play by the second

mover (level 1). However, the second mover is not truthful but rather plays backwards induction. After the first mover realizes he was wrong about the second mover, he updates his belief that the second mover plays backwards induction and from that point on the path coincides with backwards induction.

5. *Social preferences* ("soc" for short) - Play paths in which there is a departure from any of the previous categories solely because some player i removed j's top pick and j retaliated (by removing i's top pick). Everywhere else in the path the players either follow backwards induction or any of the other categories.

6. *Other* - play paths that do not belong to any of the previous categories.

For some payoff tables these six categories may overlap. For example, in some profiles truthful play and/or level 10 may coincide with a SPE play path. In light of this, we classified any play path that could be explained by a pair of SPE strategies as SPE play. Thus, all play paths that were classified into one of the other categories could not be induced by any pair of SPE strategies. In addition, we ordered the categories such that each category includes only paths that are not included in any of the preceding categories (e.g., the fifth category, *soc*, includes only paths that are not included in the first four categories). Consequently, for some payoff tables, one or more of these categories may be empty. Table 7 displays the mean proportions of play paths across the different categories for each payoff profile.

For p_1 we find that about 84% of pairs played a SPE path, which results in c (91% when the profile is played in the last ten rounds). Since this is also the observed proportion of c outcomes, it follows that whenever c was chosen, it was because the matched paired followed a SPE play path. The three most frequent categories for p_2 are SPE (38% of pairs), level $1s$ (19%) and *soc* (12%). Some of the play paths that ended with the dominated alternative c belong to the latter category. In these play paths the first mover removes the top alternative of his opponent (perhaps because the first mover believes his opponent is naïve), but is then surprised when his own top alternative

Profile	Mean	truthful	spe	level10	level1S	soc	other
	1-10		0.77				0.23
1	31-40		0.91				0.09
	Pooled		**0.84**				**0.16**
	11-20	0.10	0.25	0.15	0.21	0.13	0.16
2	21-30	0.07	0.51	0.03	0.18	0.12	0.08
	Pooled	**0.09**	**0.38**	**0.09**	**0.19**	**0.12**	**0.12**
	11-20		0.77	0.05	0.01		0.17
3	21-30		0.84	0.03	0.01		0.12
	Pooled		**0.81**	**0.04**	**0.01**		**0.15**
	1-10		0.50	0.05	0.02	0.03	0.41
4	31-40		0.49	0.21	0.04	0.07	0.19
	Pooled		**0.49**	**0.13**	**0.03**	**0.05**	**0.30**

Table 7: Mean proportions in each play-path category of AS

is removed by the opponent. From that point on, the play path coincides with the path induced by backward induction. When p_3 is played, 81% of the pairs follow the SPE path. This means that virtually all pairs that ended with the unique SPE outcome, b, indeed played the equilibrium path. This high percentage may be explained by the fact that truthful play is a SPE in this case. Finally, for p_4, about half of the pairs played an equilibrium path regardless of whether the profile was played in the first or last ten rounds. When p_4 was played in the last ten rounds about 21% of the non-SPE paths are consistent with level 10. The paths in the "other" category (41% in the first ten rounds and only 19% in the last ten rounds) do not have any salient, systematic pattern and may be the result of experimentation or mistakes by one of the players.[26]

6.3 Voting by Alternating Offers and Vetoes

6.3.1 Outcomes

For p_1 it is clear that most pairs selected the unique SPE outcome c, and its frequency increases when the profile is played in the last ten rounds (significant increase at the 1% level). Similarly, when p_2 is played, almost all pairs (95%) ended up selecting one of the SPE outcomes, a (51%) or b (44%), and the

[26]With regards to learning, the proportions of both SPE and truthful paths increase (at the 1% level) when p_1 is played in the first ten rounds versus the last ten rounds. When p_2 is played, the proportion of SPE paths increase (at the 1% level), while that of truthful remains unchanged. The proportion of SPE paths appears to be stable for p_3 and p_4, while the proportion of truthful paths decreases at the 5% level.

Profile	Mean	a	b	c	d	e
	1-10	0.04	0.05	0.81	0.06	0.04
1	31-40	0.00	0.04	0.92	0.05	0.00
	Pooled	**0.02**	**0.04**	**0.86**	**0.05**	**0.02**
	11-20	0.47	0.48	0.03	0.02	0.01
2	21-30	0.56	0.39	0.03	0.02	0.00
	Pooled	**0.51**	**0.44**	**0.03**	**0.02**	**0.00**
	11-20	0.19	0.71	0.09	0.01	0.00
3	21-30	0.06	0.84	0.07	0.01	0.02
	Pooled	**0.12**	**0.77**	**0.08**	**0.01**	**0.01**
	1-10	0.46	0.09	0.40	0.03	0.02
4	31-40	0.47	0.05	0.29	0.06	0.13
	Pooled	**0.46**	**0.07**	**0.34**	**0.05**	**0.07**

Table 8: Mean proportions of outcomes in VOAV

difference in their proportion is not statistically significant at the 10% level. For p_3, the vast majority of matched pairs (about 77%) ended up choosing the unique SPE outcome, b. Its proportion increases when the profile is played in rounds 21-30 as opposed to rounds 11-20 (significant increase at the 1% level).[27]

When payoffs are given by p_4 there is a unique SPE depending on which player moves first. It is a if player 1 moves first, and it is c if player 2 moves first, and the theoretical prediction is that a and c will be chosen and with equal proportion. When player 1 starts, about 59% of the paths end with a and 24% end with c, and when player 2 starts, about 47% of the paths end with c while 32% end with a. Thus, about 53% of the paths end with with the SPE outcome, but the proportion of a in the data is significantly higher than that of c at the 1% level. Comparing the proportion of outcomes between rounds $1-10$ and rounds $31-40$ yields the following observations. When player 1 begins, the proportion of a increases from 49% to 68%, while the proportion of c decreases from 35% to 14%. When player 2 begins, the proportion of a drops from 42% to 20%, the proportion of c barely changes (about 47%) but the proportion of e (2's top outcome) rises from 2% to 25%. Thus, while there are significant changes (in the direction of the equilibrium prediction) in the proportion of a regardless of who moves first, changes in c occur only when

[27]In this payoff table the two players disagree on whether a or c is the top element, but both agree that b is ranked in between. The mean proportions of a and c in the data are 12% and 8%, respectively. The frequency of a declines when the profile is played in the third block of ten rounds (significant change at the 1% level), while the proportion of c is stable across blocks.

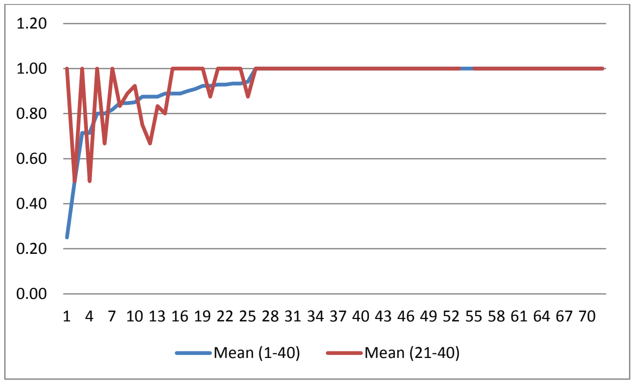

Figure 2: Backwards induction in VOAV - Mean proportion of times a player's action was consistent with backwards induction when only three items remained.

player 1 moves first.

6.3.2 Strategies

As in the previous treatment, we examined whether subjects behaved according to backward induction when only three options were left. Consider a player who needs to accept or reject an option when only three remain. If he rejects, he expects his opponent to select his highest ranked option of the two that remain. Hence, he compares that to the current proposal and optimally chooses either accept or reject (the counteroffer is irrelevant). Figure 2 displays the proportion of subjects who behaved according to this reasoning (the horizontal and vertical axes are as in Figure 1). As in the the case of AS, most subjects exhibited this minimal level of backward induction. In addition, it seems that a higher fraction of subjects exhibited this ability in the last twenty rounds of the game.

We use the same classification of play paths as in our analysis of AS. Truthful play here means that a player is not strategic: he always proposes his top ranked option (among the remaining ones), and accepts only this option. Level 10 and level $1s$ are defined in a manner that is similar to alternate strikes. The *soc* category here includes play paths with either one of the following properties: (*soc*1) the first mover proposes his top pick which gets rejected,

Profile	Mean	truthful	spe	level10	level1S	soc	other
	1-10		0.81	0.01	0.02		0.15
1	31-40		0.91	0.00	0.01		0.08
	Pooled		**0.86**	**0.01**	**0.01**		**0.12**
	11-20	0.07	0.76			0.01	0.17
2	21-30	0.04	0.77			0.04	0.15
	Pooled	**0.05**	**0.77**			**0.02**	**0.16**
	11-20		0.68	0.04		0.01	0.27
3	21-30		0.78	0.02		0.00	0.20
	Pooled		**0.73**	**0.03**		**0.00**	**0.24**
	1-10	0.02	0.27	0.03	0.22	0.05	0.42
4	31-40	0.00	0.43	0.01	0.11	0.04	0.42
	Pooled	**0.01**	**0.35**	**0.02**	**0.17**	**0.04**	**0.42**

Table 9: Mean proportions in each play-path category of VOAV

and he retaliates by rejecting the second mover's top pick, or (*soc*2) the first mover proposes his opponent's top choice. Table 9 displays the distribution of the mean proportions of play paths (calculated over all four sessions) across the different categories for each payoff table.

When payoffs were given by p_1, about 86% of matched pairs played along the SPE path, which roughly equals the frequency of the unique SPE outcome, c. When p_2 and p_3 were played, about three quarters of pairs followed a SPE path, and the majority of the remaining pairs played an unclassified path. Since in p_3 about 77% of pairs ended up with the unique SPE outcome, almost all of these pairs played a SPE path. Finally, when payoffs were given by p_4, about 35% of the pairs played a SPE path, 17% played level 1s, and almost all the remaining pairs (about 42%) followed an unclassified path.[28] This means that of the 53% of pairs that selected an SPE *outcome*, only a third did so without following the equilibrium path of play. Finally, there is some evidence that, for profiles p_1, p_3, and p_4, subjects tended to play a SPE path more frequently as they gained experience.

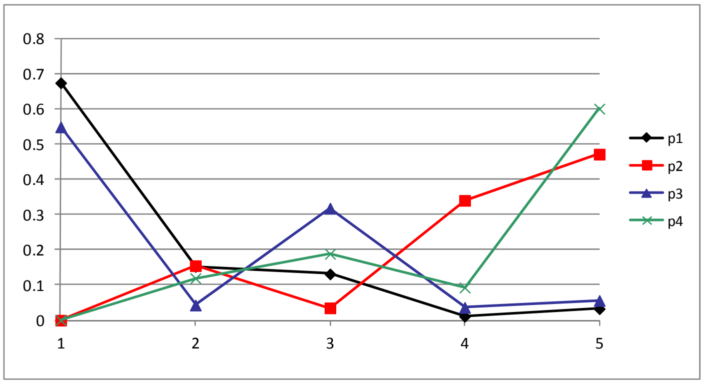

Figure 3: Distribution over ending times in VOAV (means taken over aggregated data)

6.3.3 Duration

The VAOV scheme has the appealing feature that it may end in relatively few rounds (in contrast to the AS procedure that requires four rounds of elimination). It is therefore interesting to examine the distribution of ending times of all the SPE play paths. In particular, we are interested in knowing the proportion of pairs who coordinated on the shortest equilibrium play path. This information is displayed Figure 3 (the mean proportions in the figure were computed using the aggregated data).

This figure suggests that whenever the equilibrium outcome may be perceived as a compromise (either middle ranked option as in profile 1 or second best option in profile 3), there is a relatively high likelihood that the game will end after the first proposal. However, when there is no scope for compromise (as when the first mover gets his top choice), there is a relatively high likelihood that the game will proceed until the last period.

[28] Among the 42%of pairs following other paths, around 30% of these can be categorized as "trembles by the second mover". These are cases in which the first mover begins with an action that coincides with either truthful play, SPE or level 10; the second mover "trembles" and chooses some action (possibly by mistake), which cannot be explained by the previous categories; when the first mover plays again the two players carry out backwards induction.

Profile	Mean	a	b	c	d	e
	1-10	0.05	0.06	0.79	0.07	0.04
1	31-40	0.03	0.03	0.87	0.05	0.03
	Pooled	**0.04**	**0.04**	**0.83**	**0.06**	**0.03**
	11-20	0.50	0.48	0.03	0.00	0.00
2	21-30	0.46	0.51	0.01	0.02	0.00
	Pooled	**0.48**	**0.50**	**0.02**	**0.01**	**0.00**
	11-20	0.25	0.31	0.35	0.10	0.00
3	21-30	0.41	0.25	0.23	0.10	0.02
	Pooled	**0.32**	**0.28**	**0.29**	**0.10**	**0.01**
	1-10	0.40	0.04	0.45	0.06	0.07
4	31-40	0.48	0.02	0.46	0.04	0.00
	Pooled	**0.45**	**0.03**	**0.45**	**0.05**	**0.03**

Table 10: Mean proportions of outcomes in shortlisting mechanism

6.4 Shortlisting

6.4.1 Outcomes

Recall that the equilibrium outcomes are c for p_1, both a and b for profile 2, and both a and c for p_3 and p_4. Most of the observed outcomes in profiles p_1, p_2 and p_4 coincide with the theoretical prediction. In p_3, the second ranked outcome for both players is selected about 30% of the time. This may reflect some sense of fairness on part of the subjects since the equilibrium strategy of the first mover requires him to propose a triplet containing the bottom three options of his opponent (but containing his top pick).[29]

6.4.2 Strategies

Turning next to the subjects' strategies, we look at the first movers' proposals. Since most second movers selected their preferred option from the triplet offered to them, the first mover's selection is indicative of whether or not subjects followed an equilibrium path. For each preference profile, and for each round of play, we computed the distribution over the possible triplets that player 1 could have offered. When the players' preferences are completely opposed (p_1), then in equilibrium the first mover would propose a triplet containing the

[29] There is little evidence of learning across preference profiles. There is a slight but significant rise in the frequency of c in p_1 (5% significance), and a larger increase in the proportion of a in p_3 (1% significance). With the exception of these, there are no significant changes in the proportions of outcomes when preference profiles are played later in the experiment.

bottom three elements of his opponent ($\{a, b, c\}$ for player 1 and $\{c, d, e\}$ for player 2). These triplets account for about 90% of the first period proposals. Proposing the bottom three elements of the opponent is also the equilibrium strategy when the players disagree on the first and third best outcome but agree on the second best (p_3). However, in contrast to p_1, here there is scope for a compromise in the form of the second-best alternative. This may explain why the proportion of equilibrium first period proposals (which exclude the compromise) is only about 56%. When the players disagree only on the top two alternatives (p_2), the first mover's equilibrium proposal includes his top option but excludes his opponent's top option, and 78% of proposals have this feature.[30] Finally, in the fourth profile, the equilibrium offers are $\{a, b, d\}$ and $\{c, d, e\}$, and the frequency of these offers is about 72%.[31]

A complementary approach to understanding subjects' strategic behavior is to categorize the observed play paths. We consider three natural categories: *truthful* (proposing one's top three elements and choosing the top ranked option), *SPE* (note that this also includes a play path in which the first mover is level 1 and the second mover is level zero) and *social preferences*, which includes play paths that satisfy one of the following. (*soc*1) The proposer offers the equilibrium outcome (which is *not* the responder's top choice) together with either dominated options or the responder's bottom two options; the responder "punishes" the proposer by selecting an option, which is worse for the proposer than the equilibrium outcome. (*soc*2) The first mover, i, proposes a set that includes any of the following: the top pick of the second mover, player j, a "compromise" (the common second best) or the outcome that would have been an equilibrium had j moved first; the second mover, j, does *not* pick a

[30] There is an interesting learning trend in this profile. When players just start playing this profile, about 50% of the first movers proposes the top three elements of both players. As the players gain experience, they gradually move away from this proposal until finally the proportion of this proposal falls below 10%.

[31] When this profile is played in the first ten rounds, subjects in the role of player 1 also offer $\{a, b, c\}$ (about 23%), which includes the top three options of player 1 and the middle three options of player 2. One reason for this may be that this offer is perceived as fair (the top choice for player 2 is excluded since it is the worst option for player 1). However, when the profile is played in rounds 31-40, then this offer is no longer made in the last five rounds.

Profile	Mean	spe	truthful	soc	other
	11-20	0.65	0.28	0.07	0.01
2	21-30	0.83	0.07	0.09	0.01
	Pooled	**0.74**	**0.17**	**0.08**	**0.01**
	11-20	0.43	0.13	0.44	0.00
3	21-30	0.49	0.09	0.43	0.00
	Pooled	**0.46**	**0.11**	**0.44**	**0.00**
	1-10	0.46	0.29	0.21	0.04
4	31-40	0.84	0.08	0.07	0.01
	Pooled	**0.67**	**0.18**	**0.13**	**0.03**

Table 11: Mean proportions in each play-path category of shortlisting mechanism

dominated option.[32] We label any play path as "other" if it does not fit any of these three categories (truthful, SPE or soc).

The SPE and truthful categories coincide only when the players' preferences are completely opposed (p_1). In that case, the mean proportion of SPE/truthful play paths is 78% in the first ten rounds and 87% in the last ten rounds. This difference in proportions is statistically significant at the 1% level.

Table 11 displays the mean proportions across these categories for profiles p_2 to p_4. As shown, social preferences appear to play an important role in explaining deviations from equilibrium play (especially since the first mover's action is relatively simple compared to the previous sequential schemes). This is reflected in p_3 where we observe the highest percentage (44%) of social preferences play paths. Truthful play for this profile calls for the first mover to propose his top three elements. This is somewhat naïve as the second mover will simply pick his own top choice, which will leave the first mover with his third-best choice. Indeed, only 11% of pairs exhibit this path. The equilibrium path may be perceived as somewhat unfair as it calls the first mover to propose the bottom three options of the second mover. This may explain why only 46% of participants exhibit this path. A more fair proposal for the first mover, which is not as naïve as truthful play, is to propose his own top choice, the compromise option b (which is second-best to both parties) and one of his (and the other party's) bottom two choices. The majority of the *soc* play paths fall into this category. The *soc* category also includes instances

[32]The k-level framework is less appropriate for this mechanism as the second player simply chooses the best alternative from a set of three.

in which the first mover followed the equilibrium by proposing his top ranked choice, but the second mover choose dominated outcomes possibly to ”punish” them for being unfair.[33]

We conjecture that social preferences have a much lower impact in real-world arbitration cases. Hence, if shortlisting were to be adopted by arbitration agencies, we expect the behavior to be more in line with equilibrium.

Finally, interesting trends emerge in these data, with the likelihood of equilibrium play higher in later rounds than in earlier rounds for both profiles p_2 and p_4.

6.5 Comparing the mechanisms

In this section we compare the “performance” of the different mechanisms along two dimensions. The first dimension focuses on outcomes and the associated payoffs. That is, we examine to what extent the resulting outcomes were “socially desirable”. The second dimension focuses on strategies. That is, we examine to what extent the participants followed an equilibrium path. For this dimension, we focus only on the three sequential mechanisms, where the set of equilibrium outcomes coincide with those implemented by the social choice function.[34]

To carry out this comparison consider the following incomplete binary relation defined over mechanisms. For a given measure of performance, mechanism x is said to weakly dominate mechanism y if two conditions are satisfied: (i) there exists at least one profile for which x outperforms y and this difference is significant at a $p < 10\%$, and (ii) there exists no preference profile for which y outperforms x and this difference is statistically significant at any $p < 15\%$. All the measures discussed in this section were computed for each preference profile as the average over all rounds in which the profile was played.

[33]Such apparent punishments by the second mover explain the gap between the percentage of cases in which the first-mover played an equilibrium strategy and the proportion of overall equilibrium plays (which require the second-mover to also adhere to the equilibrium).

[34]As shown in Section 3, equilibrium outcomes do not necessarily coincide with those implemented by the social choice function for the rank-veto mechanism.

Profile 2	t2-t1	t3-t1	t4-t1	t3-t2	t4-t2	t4-t3
Coef.	0.087	-0.036	-0.056	-0.123	-0.143	-0.020
Std. Err.	0.021	0.017	0.015	0.019	0.018	0.013
t	4.170	-2.150	-3.650	-6.450	-7.960	-1.600
P>|t|	0.000	0.033	0.000	0.000	0.000	0.112
Profile 3	**t2-t1**	**t3-t1**	**t4-t1**	**t3-t2**	**t4-t2**	**t4-t3**
Coef.	-0.082	-0.098	-0.013	-0.016	0.069	0.085
Std. Err.	0.019	0.018	0.023	0.010	0.017	0.016
t	-4.320	-5.360	-0.570	-1.620	4.050	5.240
P>|t|	0.000	0.000	0.567	0.108	0.000	0.000
Profile 4	**t2-t1**	**t3-t1**	**t4-t1**	**t3-t2**	**t4-t2**	**t4-t3**
Coef.	-0.052	-0.070	-0.119	-0.018	-0.067	-0.049
Std. Err.	0.024	0.024	0.023	0.019	0.018	0.019
t	-2.140	-2.870	-5.080	-0.940	-3.700	-2.640
P>|t|	0.034	0.005	0.000	0.347	0.000	0.009
Profile 4*	**t2-t1**	**t3-t1**	**t4-t1**	**t3-t2**	**t4-t2**	**t4-t3**
Coef.	-0.044	-0.026	-0.112	0.019	-0.068	-0.087
Std. Err.	0.023	0.027	0.023	0.023	0.018	0.023
t	-1.950	-0.950	-4.860	0.800	-3.720	-3.700
P>|t|	0.053	0.342	0.000	0.423	0.000	0.000

Table 12: Comparing proportion of dominated and unfair outcomes across treatments

We propose two measures for comparing the performance of the mechanisms in terms of outcomes. The first measure is the mean frequency of alternatives that are either Pareto dominated or "unfair" in the sense of being worse than some player's median choice for a given preference profile. We apply these measures only to the profiles with dominated alternatives, namely p_2, p_3 and p_4. Note that in the second and third profiles, the set of Pareto dominated alternatives contains the set of unfair outcomes. Only in the fourth profile there is an outcomes, e, which is undominated and yet unfair. According to this measure, mechanism x outperforms mechanism y if the mean proportion of dominated/unfair outcomes in x is lower than under y. Table 12 compares the four mechanisms according to this measure for each preference profile. We have included two measures for profile 4, one which looks only at dominated alternative, and one which looks at both dominated and unfair. Note that we have excluded the first preference profile where there are no dominated options.

From Table 12 it follows that shortlisting weakly dominates the rank-veto procedure at the 1% level; the VAOV scheme weakly dominates the AS procedure at the 1% level, and in all profiles but the third one, shortlisting outperforms both the AS and the VAOV procedure at the 1% level. Hence, there

Profile 2	t2-t1	t3-t1	t4-t1	t3-t2	t4-t2	t4-t3
Coef.	-0.087	0.043	0.069	0.130	0.156	0.026
Std. Err.	0.027	0.022	0.021	0.024	0.023	0.017
t	-3.230	1.930	3.300	5.350	6.730	1.510
P>|t|	0.002	0.055	0.001	0.000	0.000	0.132
Profile 3	**t2-t1**	**t3-t1**	**t4-t1**	**t3-t2**	**t4-t2**	**t4-t3**
Coef.	0.200	0.226	0.059	0.026	-0.142	-0.167
Std. Err.	0.046	0.044	0.052	0.022	0.036	0.034
t	4.400	5.090	1.120	1.140	-3.990	-4.910
P>|t|	0.000	0.000	0.263	0.255	0.000	0.000
Profile 4	**t2-t1**	**t3-t1**	**t4-t1**	**t3-t2**	**t4-t2**	**t4-t3**
Coef.	0.084	0.043	0.075	-0.041	-0.009	0.033
Std. Err.	0.036	0.038	0.039	0.032	0.033	0.036
t	2.350	1.110	1.930	-1.290	-0.270	0.910
P>|t|	0.020	0.268	0.055	0.199	0.790	0.363

Table 13: Comparing efficiency measure (ratio of excess surplus above random choice to maximal surplus) across treatments

is a weak sense in which shortlisting performs best according to the above measure.

An alternative measure of the efficiency of a mechanism compares the total realized surplus with the maximal achievable surplus. Note that choosing an option at random with equal probability generates a surplus of \$1.[35] We therefore propose to consider random choice as a reference point and to look at the ratio of the "excess surplus" - the difference between the realized surplus and the surplus from random choice - to the maximal achievable surplus. Table 13 displays the pair-wise comparisons according to this measure.[36]

As shown, the VAOV scheme appears to perform best according to this alternative measure. Both the VAOV scheme and shortlisting weakly dominate the rank-veto mechanism at the 10% level. But the VAOV scheme also weakly dominates both the AS procedure and shortlisting at the 1% level.

Turning to the strategy dimension, table 14 displays the pair-wise comparisons according to the fraction of times the equilibrium path was played. We have excluded here the first preference profile for which all procedures performed equally well.

As shown, shortlisting dominates the AS scheme at the 1% level in p_2 and

[35]In the third pair of preferences, for instance, picking a, b, or c systematically leads to the maximal surplus, \$1.5. Picking d leads to 0.25+0.25=\$0.5, while picking e leads to zero. So the 'surplus from random choice' is (1.5+1.5+1.5+0.5+0)/5=1

[36]Note that we have excluded the first preference profile where any outcome leads to the maximal surplus.

Profile 2	t3-t2	t4-t2	t4-t3
Coef.	0.388	0.372	-0.017
Std. Err.	0.038	0.040	0.033
t	10.130	9.240	-0.500
P>|t|	0.000	0.000	0.616
Profile 3	**t3-t2**	**t4-t2**	**t4-t3**
Coef.	-0.078	-0.354	-0.276
Std. Err.	0.030	0.038	0.037
t	-2.600	-9.250	-7.380
P>|t|	0.010	0.000	0.000
Profile 4	**t3-t2**	**t4-t2**	**t4-t3**
Coef.	-0.150	0.174	0.324
Std. Err.	0.036	0.037	0.041
t	-4.130	4.720	7.830
P>|t|	0.000	0.000	0.000

Table 14: Comparing mean proportion of equilibrium play paths across treatments

p_3. Note that this observation is common to all three criteria for comparison (Tables $12-14$). As discussed in the previous subsection, social preferences lead deviations from equilibrium play in the shortlisting procedure, whereas they induce adherence to equilibrium in the AS scheme (where truthfulness is an equilibrium). However, shortlisting and the VAOV scheme are not comparable: shortlisting outperforms the VAOV at the 1% level in p_4, the VAOV outperforms shortlisting at the 1% level in p_3 but the difference between the two is not statistically significant in p_2.

7 Concluding remarks

This paper takes an implementation-theoretic approach to the problem of assigning a public good, namely an arbitrator, to two parties with symmetric information. As a first step, we prove that to have a mechanism with "socially desirable" properties, one must consider sequential mechanisms and alternative SCRs to the one induced by truthtelling in the most common selection procedure, the simultaneous rank-veto mechanism. After presenting empirical evidence that supports the hypothesis of strategic behavior in the rank-veto mechanism, we test alternative selection procedures in the lab. The experimental analysis yields three key results. First, a large fraction of players followed strategic behavior, suggesting that the rank-veto procedure may suffer from the deficiencies outlined in the theoretical section. Second, we find substantial evidence that players follow backwards induction under the sequential mechanisms. Third, we find that the VAOV scheme and shortlisting appear to perform better than the other procedures in spite of the fact that neither of these mechanisms are used in practice.

While our results are presented in the context of arbitrator selection, they potentially may be extended to other situations in which a collective of individuals with symmetric information need to agree on a public good (i.e., an outcome that affects the payoffs of the participants). Examples may include hiring decisions, choosing a set of employees to promote, selecting jury members and deciding on the composition of some committee. Our paper suggests that it may be valuable to study these situations from an implementation-theoretic approach: start by identifying "reasonable" SCRs for the problem at hand; ask whether prevalent procedures implement in theory any of these SCRs; study whether participants in such mechanisms tend to behave according to theory; explore alternative mechanisms that "perform well" both theoretically and behaviorally.

References

Abdulkadiroglu, A., P. A. Pathak, and A. E. Roth, 2005. The New York City High School. *American Economic Review (Papers and Proceedings)* **95**, 364-367.

Abdulkadiroglu, A., P. A. Pathak, A E. Roth, and T. Sönmez, 2005. The Boston Public School. *American Economic Review (Papers and Proceedings)* **95**, 368-371.

Anbarci, N., 2006. "Finite Alternating-Move Arbitration Schemes and the Equal Area Solution." *Theory and Decision* **61**, 21-50.

Ashenfelter, O., and G. B. Dahl, 2005. Strategic Bargaining Behavior, Self-Serving Biases, and the Role of Expert Agents: An Empirical Study of Final-Offer Arbitration. NBER Working Papers 11189.

Binmore, K., J. McCarthy, G. Ponti, L. Samuelson, and A. Shaked,

2002. A Backward Induction Experiment. *Journal of Economic Theory* **104**, Pages 48-88.

Bloom, D. E., and C. L. Cavanagh, 1986a. An Analysis of the Selection of Arbitrators. *American Economic Review* **76**, 408-422.

Bloom, D. E., and C. L. Cavanagh, 1986b. An Analysis of Alternative Mechanisms for Selecting Arbitrators. Harvard Institute of Economic Research Discussion Paper 1224.

Brams, S. J. and D. M. Kilgour, 2001. "Fallback Bargaining." *Group Decision and Negotiation* **10**, 287-316.

Crawford, V. P., M. Costa-Gomes, and N. Iriberri, 2010. "Strategic Thinking". Mimeo, UCSD.

Degana, A., and A. Merlo, 2009. Do voters vote ideologically? *Journal of Economic Theory* **144**, 1868-1894.

Forsythe, R., R. B. Myerson, T. A. Rietz, and R. J. Weber, 1993. An Experiment on Coordination in Multi-Candidate Elections: The Importance of Polls and Election Histories. *Social Choice and Welfare* **10**, 223-247.

Forsythe, R., R. B. Myerson, T. A. Rietz, and R. J. Weber, 1996. An Experimental Study of Voting Rules and Polls in Three-Candidate Elections. *International Journal of Game Theory* **25**, 355-383.

Hurwicz, L, and D. Schmeidler, 1978. Construction of Outcome Functions Guaranteeing Existence and Pareto Optimality of Nash Equilibria. *Econometrica* **46**, 1447-1474.

Hurwicz, L., and M. R. Sertel. 1997. "Designing Mechanisms, in particular for Electoral Systems: The Majoritarian Compromise.' ' Mimeo.

Kawai, K., and Y. Watanabe, 2010. Inferring Strategic Voting. Mimeo, Northwestern University.

Kibris, O. and M. R. Sertel, 2007. "Bargaining Over a Finite Set of Alternatives." *Social Choice and Welfare* **28**, 421-437.

Klement, A. and Z. Neeman, 2011. "Private Selection and Arbitrator Impartiality." Mimeo, Tel-Aviv University.

Levitt, S. D., J. A. List, and S. E. Sadoff, 2010. Checkmate: Exploring Backward Induction Among Chess Players. NBER Working Paper 15610.

Moore, J., 1992. Implementation, Contracts, and Renegotiation in Environments with Complete Information. In J.J. Laffont (ed.), Advances in Economic Theory, Vol. 1, Cambridge: Cambridge University Press, 182-282.

Roth, A. E., 1984. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy* **92**, 991-1016.

Roth, A. E., 2007. The Art of Designing Markets. *Harvard Business Review*, October Issue, 118-126.

Sprumont, Y., 1993. "Intermediate Preferences and Rawlsian Arbitration Rules." *Social Choice and Welfare* **10**, 1-15.

Appendix 1: Instructions and quiz for Shortlisting treatment

The Game:

This is an experiment in interactive decision-making. By participating in this experiment, you will win a show-up bonus of $10 plus a prize that depends upon the choices that you and other participants make.

The experiment consists of 40 rounds. In each round the computer will randomly match you to another participant and both of you will play the game described below. The outcome of this game is a selection of a single option from a list of five, labeled a,b,c,d,e.

In each round, each of the five options will be assigned a dollar value. One option will be worth $1.00, a second option will be worth $0.75, a third option will be worth $0.50, a fourth option will be worth $0.25 and a fifth will be worth $0.00. At the beginning of each round, you will be told what dollar value you assign to each of the options, as well as what dollar value the other player assigns to each of the options. The table below shows a possible configuration of values for you and the other player (the values in the experiment itself) :

	You	Other Player
$1.00	d	c
$0.75	b	a
$0.50	a	e
$0.25	e	b
$0.00	c	d

Similarly, the participant you are matched with will also be shown the values that both he and you assign to each option. In other words, both you and the participant you are matched with will see the same table as the one shown above.

The 40 rounds will be divided into four stages consisting of 10 rounds each (such that stage 1 consists of rounds 1-10, stage 2 consists of rounds 11-20, stage 3 consists of rounds 21-30 and stage 4 consists of rounds 31-40). In each stage, half of the participants will be randomly chosen to belong to group A and the other half will belong to group B. The participants belonging to the same group will have the same assignment of money to options. In each of the 10 rounds of a stage, the participants of one group will be randomly matched to the participants in the other group. Thus, in each stage of 10 rounds, the values you assign to each option and the values of your matched participant will remain unchanged.

At the beginning of each stage you will be shown a message that announces the start of a new stage. This message will alert you to the fact that the values of the options for you and the other player may be different from what they were in the previous stage. As in every round, these values will be displayed on the screen.

Your total payoff in the entire experiment will equal the sum of payoffs across all 40 rounds plus a show-up bonus of $10.

Your payoff in each round will be displayed on the top right of the screen.

We now describe the rules of the game that you will play in each of the 40 rounds. These rules determine which option is selected at the end of the round.

- In each round, one of the two participants who are matched to play the game, is randomly selected to be Player 1 (the other participant is then Player 2). This means that each of the two participants has an equal chance of being selected as Player 1. This also means that each participant may be in the role of

Player 1 on some rounds and in the role of Player 2 on other rounds.

- The game proceeds in two steps
    - Step 1: Player 1 moves first and selects a shortlist of three distinct options out of a,b,c,d,e.
    - Step 2: Player 2 is informed of Player 1's shortlist, and chooses the final option out of it.

To illustrate these rules, here is a simple example.

Suppose that in a particular round with payoffs as listed in the table above, you have been selected to be Player 1, while the other participant you are matched with had been selected to be Player 2. Suppose that the following actions have been chosen

- Player 1 (You) moves first and selects the shortlist a, c, d.
- Player 2 (the other participant you are matched with) moves second and chooses c out of a, c, d.

This round thus ends with c being selected. You thus receive a payoff of $0 for this round, while the participant you are matched with receives a payoff of $1.

Quiz:
To confirm whether you understood the rules of the game, please answer the following question.
Assume that the values you assign to each option are as follows:

	You
$1.00	c
$0.75	e
$0.50	d
$0.25	b
$0.00	a

Question 1. Suppose that in a particular round, players choose the following actions.

- Player 1 moves first and selects the shortlist a, c, e.
- Player 2 moves second and selects c out of the shortlist a, c, e.

(**a**) Which option will be selected?

- ○ a
- ○ b
- ○ c
- ○ d
- ○ e

(**b**) What will your payoff be?

- ○ $1.00
- ○ $0.75
- ○ $0.50
- ○ $0.25
- ○ $0.00

Question 2. Suppose that in a particular round, the participant you are matched was selected to be Player 1, and selected the shortlist b, d, e.

(**a**) Suppose that you selected the option d out of the shortlist. What will your payoff be?

- $1.00
- $0.75
- $0.50
- $0.25
- $0.00

(**b**) Suppose that you selected the option e out of the shortlist. What will your payoff be?

- $1.00
- $0.75
- $0.50
- $0.25
- $0.00

Click to send your answers to the quiz

Appendix 2: Veto matrices of simultaneous mechanism

Player 1 \ Player 2	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	
AB											**0.00%**
AC											**0.00%**
AD						0.29%					**0.29%**
AE											**0.00%**
BC											**0.00%**
BD	0.29%		1.71%			4.29%			0.29%	1.43%	**8.00%**
BE						0.57%					**0.57%**
CD	0.29%					0.86%		0.29%		0.29%	**1.71%**
CE	6.00%		8.29%	0.29%		24.00%			1.43%	0.57%	**40.57%**
DE	12.00%	0.29%	8.57%	0.57%	0.29%	24.29%	0.29%	0.57%		2.00%	**48.86%**
	18.57%	**0.29%**	**18.57%**	**0.86%**	**0.29%**	**54.29%**	**0.29%**	**0.86%**	**1.71%**	**4.29%**	**100.00%**

Profile p_1

Player 2

Player 1	**AB**	**AC**	**AD**	**AE**	**BC**	**BD**	**BE**	**CD**	**CE**	**DE**	
AB										0.57%	**0.57%**
AC											**0.00%**
AD											**0.00%**
AE											**0.00%**
BC		0.57%							0.57%	2.29%	**3.43%**
BD		0.29%	0.29%							1.14%	**1.71%**
BE		1.43%	0.86%	1.14%				0.86%	0.29%	4.86%	**9.43%**
CD			0.29%	0.29%					1.14%	2.00%	**3.71%**
CE			0.29%	1.14%						2.00%	**3.43%**
DE		4.86%	8.86%	12.57%	0.57%	0.29%		1.14%	4.29%	45.14%	**77.71%**
	0.00%	**7.14%**	**10.57%**	**15.14%**	**0.57%**	**0.29%**	**0.00%**	**2.00%**	**6.29%**	**58.00%**	**100.00%**

Profile p_2

Player 2

Player 1	**AB**	**AC**	**AD**	**AE**	**BC**	**BD**	**BE**	**CD**	**CE**	**DE**	
AB											**0.00%**
AC											**0.00%**
AD											**0.00%**
AE											**0.00%**
BC	1.71%		2.29%	1.43%						6.86%	**12.29%**
BD											**0.00%**
BE											**0.00%**
CD	0.86%		2.57%	0.57%						6.57%	**10.57%**
CE	4.86%		1.43%	1.14%						8.29%	**15.71%**
DE	7.71%		5.43%	9.71%						38.57%	**61.43%**
	15.14%	**0.00%**	**11.71%**	**12.86%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**60.29%**	**100.00%**

Profile p_3

Player 2

Player 1	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE	
AB											**0.00%**
AC											**0.00%**
AD						0.29%					**0.29%**
AE											**0.00%**
BC											**0.00%**
BD	0.29%		1.71%			4.29%			0.29%	1.43%	**8.00%**
BE						0.57%					**0.57%**
CD	0.29%					0.86%		0.29%		0.29%	**1.71%**
CE	6.00%		8.29%	0.29%		24.00%			1.43%	0.57%	**40.57%**
DE	12.00%	0.29%	8.57%	0.57%	0.29%	24.29%	0.29%	0.57%		2.00%	**48.86%**
	18.57%	**0.29%**	**18.57%**	**0.86%**	**0.29%**	**54.29%**	**0.29%**	**0.86%**	**1.71%**	**4.29%**	**100.00%**

Profile p_4

Appendix 3: Examples of agencies and the mechanisms they use for selecting arbitrators

Procedure	Agency	Link	Section
Rank-Veto	FMCS	http://www.fmcs.gov/internet_text_only/itemDetail.asp?categoryID=197&itemID=16959	1404.12 (c)
	AAA	http://www.adr.org/sp.asp?id=22440#R11	R-11 (b)
	AIRROC	http://www.airroc.org/drp/default.aspx?file=THE.AIRROC.DISPUTE.RESOLUTION.PROCEDURE.pdf	3
	CBAAS	https://www.cincybar.org/arbitration_Rules.asp	VI.
	Hong Kong Construction Association	http://www.hkca.com.hk/upload/files/0100015727.pdf	Appendix 3: 2.7
	State of New Hampshire	http://www.nh.gov/pelrb/forms/documents/arbitration.pdf	4
Alternate Strike	FMCS	http://www.fmcs.gov/internet_text_only/itemDetail.asp?categoryID=197&itemID=16959	1404.12 (c)
	California Code of Civil Procedure	http://law.justia.com/codes/california/2009/ccp/1299-1299.9.html	1299.4. (c)
	Virginia Acts of Assembly	http://leg1.state.va.us/cgi-bin/legp504.exe?051+ful+CHAP0356+pdf	1. B.
	SEC Union	http://www.secunion.org/CBAArticle33	33: 2 A
	Oregon Revised Statutes	http://www.leg.state.or.us/03reg/measures/sb0400.dir/sb0444.a.html	243.746
	ARIAS U.S.	http://www.arias-us.org/index.cfm?a=91	4
	Montana Code	http://data.opi.mt.gov/bills/mca/20/4/20-4-204.htm	5
	City of Sacramento Charter	http://www.qcode.us/codes/sacramento/view.php?topic=city_of_sacramento_charter-xix-603&frames=on	B)
	SUNY Brockport	http://www.brockport.edu/policies/docs/attendance_policy_appeal_process.pdf	II. A.
	City of San Luis Obispo	http://www.slocity.org/cityclerk/elections/2011/Measure%20B/2000Election/Full%20Text%20-Meas%20S.pdf	(d) (2)
	Airline Labor Dispute Resolution Act	http://railwaylaboract.com/aldra-s.1327.htm	2. (2)
	Ventura County Community College	http://www.vcccd.edu//assets/pdf/human_resources/aft_16.pdf	16.12 V: B.
	CUE Union	http://www.cueunion.org/bargaining/contract/art03.php	C. 2.
	University of Michigan	http://www.vpcomm.umich.edu/gsi-sa/contract05-08-a16.html	G. 4. c.
	State of Wisconsin	http://legis.wisconsin.gov/2009/data/AB-450.pdf	Page 3, Lines 12-16
	City of Colorado Springs	http://www.springsgov.com/units/municourt/faqjurors.htm	
	Cecil County	http://www.ccgov.org/commissioners/ordinance/2011.04.sheriffsofficeamendment.pdf	5