

NBER WORKING PAPER SERIES

TRADING VOTES FOR VOTES. A DECENTRALIZED MATCHING ALGORITHM

Alessandra Casella

Thomas Palfrey

Working Paper 21645

<http://www.nber.org/papers/w21645>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

October 2015

We thank Enrico Zanardo, Kirill Pogorelskiy and Manuel Puente for research assistance, and participants to the June 2013 Alghero (Italy) workshop on "Institutions, Individual Behavior and Economic Outcomes", the 2015 Warwick/Princeton Political Economy Conference (Venice, Italy), the 2015 Montreal Political Economy Workshop, the 2015 Princeton Workshop on Social Choice and Public Economics, the 2015 Wallis Institute Annual Conference at the University of Rochester, and to seminars at New York University, the University of Michigan, and the University of Pittsburgh for comments. We especially wish to thank Micael Castanheira, Andrew Gelman, Debraj Ray, Richard Van Veelden, and Alistair Wilson for detailed comments and suggestions. The National Science Foundation and the Gordon and Betty Moore Foundation provided financial support. Part of the research was conducted while Casella was a Straus Fellow at NYU Law School and Palfrey was a Visiting Scholar at the Russell Sage Foundation. The hospitality and financial support of both institutions are gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Alessandra Casella and Thomas Palfrey. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Trading Votes for Votes: A Decentralized Matching Algorithm

Alessandra Casella and Thomas Palfrey

NBER Working Paper No. 21645

October 2015

JEL No. C92, D7, D72

### **ABSTRACT**

Vote-trading is common practice in committees and group decision-making. Yet we know very little about its properties. Inspired by the similarity between the logic of sequential rounds of pairwise vote-trading and matching algorithms, we explore three central questions that have parallels in the matching literature: (1) Does a stable allocation of votes always exist? (2) Is it reachable through a decentralized algorithm? (3) What welfare properties does it possess? We prove that a stable allocation exists and is always reached in a finite number of trades, for any number of voters and issues, for any separable preferences, and for any rule on how trades are prioritized. Its welfare properties however are guaranteed to be desirable only under specific conditions. A laboratory experiment confirms that stability has predictive power on the vote allocation achieved via sequential pairwise trades, but lends only weak support to the dynamic algorithm itself.

Alessandra Casella

Department of Economics

Columbia University

420 West 118 Street

New York, NY 10027

and NBER

ac186@columbia.edu

Thomas Palfrey

Division of Humanities and Social Sciences

228-77 Caltech

Pasadena, CA 91125

and NBER

trp@hss.caltech.edu

# 1 Introduction

Trading support for one proposal in exchange for someone else's support of a different proposal is a common aspect of voting in committees, legislatures, and other bodies of group decision making. Whether as exchanges of favors in small informal committees or as more elaborate deals in legislatures, common sense, anecdotes, and systematic evidence, all suggest that the practice is a central component of decision-making in groups.<sup>1</sup> Yet we know very little about the properties of vote trading. Efforts at a theory were numerous and enthusiastic in the 1960's and 70's but have fizzled and disappeared in the last 40 years. John Ferejohn's words in 1974, towards the end of this wave of research, remain true today: "[W]e really know very little theoretically about vote trading. We cannot be sure about when it will occur, or how often, or what sort of bargains will be made. We don't know if it has any desirable normative or efficiency properties." (Ferejohn, 1974, p. 25)

One reason for the lack of progress is that the problem is difficult: not only does vote bartering occur without the equilibrating properties of a continuous price mechanism, not only does it cause externalities to allies and opponents of the trading parties, but each exchange triggers new profitable exchanges. If we think of the trades sequentially, as a subset of voters trade votes on a set of proposals, the default outcomes of these proposals change, generating incentives for a new round of vote trades, which will again change outcomes and trigger new trades. What is the most productive approach to modeling such a complex process?

The perspective taken in this paper is inspired by a similarity between the logic of sequential rounds of vote trading and the problem of achieving stability in sequential rounds of matching among different agents, as originally proposed by Gale and Shapley (1962).<sup>2</sup> In line with the matching literature, we explore the properties of a class of algorithms through which a sequence of decentralized pairwise vote trades are realized, and in particular we use the familiar notion of stability from that literature. An allocation of votes is stable if no pairwise-improving vote trade

---

<sup>1</sup>There is an substantial literature in political science documenting vote trading in legislatures. For example, Stratmann (1992) provides evidence of vote trading in agricultural bills in the US Congress.

<sup>2</sup>Roth and Sotomayor (1990) and Gusfield and Irving (1989) survey some of the main results, from the perspective of economic theory and computer science, respectively. This continues to be a very active area of research.

exists.<sup>3</sup> We ask whether a stable allocation of votes exists, whether the specific algorithm we construct converges to a stable allocation, and whether we can say anything about individuals' preferences over the outcomes induced by stable vote allocations.

We should note at the outset that the parallel to matching problems is imperfect. The closest analogue is to a one-sided matching problem with externalities: a single group of individuals who match in pairs but such that everyone has preferences not only over his own partner, but over the composition of all matches. Here too there is only one group—voters—and in principle everyone can match with everyone else,<sup>4</sup> and preferences are defined over the full set of matches. But in addition in our voting problem preferences evolve endogenously in response to executed trades. Trades by others can reverse a voter's status as winner or loser and affect his desire to trade, as well as his attractiveness as trading partner.

The committee we study is formed by an odd number of voters and faces several proposals, each of which may pass or fail and, after trade, is voted upon separately through majority voting. Every committee member can be in favor or opposed to any proposal and attaches some cardinal value to his preferred direction prevailing. Members' preferences are separable across proposals. A vote trade is a physical exchange of ballots. For most of our analysis and in the experiment, we allow pair-wise trading only: two voters engage in a trade if one delivers his vote to the other on one proposal, in exchange for the other's vote on a different proposal. As in the matching literature, a trading pair is said to *block* a given allocation of votes if it can be better-off under an alternative allocation that is in the pair's power to achieve, keeping fixed the votes held by the other committee members. A stable allocation is then an allocation of votes that cannot be blocked, i.e. such that no pair-wise improving trade exists.

We define a further restrictions of the possible stable vote allocations as those that are achievable from the initial vote allocation via a sequence of pairwise trades. We consider a family of trading algorithms, according to which these trades can

---

<sup>3</sup>This notion of stability has also been used in the analysis of network formation. See Jackson and Wolinsky (1996) for an early application, and Jackson and Watts (2002) for the analysis of a dynamic algorithm building stable network configurations through the creation of pairwise improving links.

<sup>4</sup>In contrast to two-sided matching problems where matches only occur between members of two separate groups—men and women, students and schools, workers and firms.

take place. In such an algorithm, an initial payoff improving exchange of votes between two voters is selected among all possible pairwise improving trades, using a specific selection rule, including random rules. This leads to a new allocation of votes, and the algorithm again selects a pairwise improving trade from the set of all pairwise improving trades. The algorithm continues until a vote allocation is reached where there are no more pairwise improving trades. The family of such algorithms is populated by considering all possible selection rules.

As remarked in Riker and Brams (1973), the requirement that a vote trade be welfare improving for both traders implies that the votes being trade must be pivotal, and we call the class of such algorithms the *Pivot Algorithms*. Our first result is that a Pivot algorithm *always* generates a stable vote allocation in a finite number of steps, for any number of voters, any number of proposals, and any configuration of (separable) preferences.

This is an interesting result, not only for its generality but also because stability—convergence to a vote allocation such that no further vote trade is profitable—was one of the two central questions of the early literature on vote trading. The literature addressed the ambitious conjecture that vote trading may offer the solution to majority cycles in the absence of a Condorcet winner. The original analysis (Park (1967)), studied non-binding agreements when voters vote on the full package of proposals (as opposed to voting separately on each proposal). Park considered only majority coalitions and concluded that the process can converge only if a Condorcet winner exists.<sup>5</sup> Riker and Brams (1973) and Ferejohn (1974) simplified the problem by considering binding agreements and proposal-by-proposal voting, as in our model. Their conclusions are ambiguous: Riker and Brams conjecture that even if stability held for pair-wise trading it would be compromised by allowing trades among larger coalitions of voters; Ferejohn suggests that a stable vote allocation may not hold even for pair-wise trading if voters are forward-looking, but does not fully specify the game structure.

The Pivot algorithm through which we model vote trades implies that voters are myopic, and if pair-wise trades only are allowed, a stable vote allocation always exists. In the second part of the paper, we allow vote-trading among coalitions

---

<sup>5</sup>The result was later echoed by other studies, for example Berholz (1973), Koehler (1975), Schwartz (1981).

of voters of any sizes. Our results confirm Riker and Brams' conjecture. With coalition-trading stability cannot be guaranteed: we construct an example with well-behaved preferences where a cycle develops and trading need never end. In addition, in our model, with proposal-by-proposal voting and binding trades, there is no logical connection between coalitional stability in vote-trading and the existence of a Condorcet winner. A coalition-stable vote allocation may exist in the absence of a Condorcet winner, and may differ from the Condorcet winner when the Condorcet winner exists.

The second conjecture at the core of the interest in vote trading in the 60's and 70's concerned not the existence of stable vote allocations but their welfare properties. It held that vote trading leads to Pareto superior outcomes because it allows the expression of the intensity of preferences. The conjecture stemmed from an early debate between Gordon Tullock and Anthony Downs<sup>6</sup> and was stated explicitly in Buchanan and Tullock (1965).<sup>7</sup> As a general result, the claim was rejected by Riker and Brams' (1973) influential "paradox of vote trading": Riker and Brams showed that if vote trading is pair-wise and binding, there are non-pathological preferences such that each pair of voters individually gains from vote trading and yet everyone strictly prefers the no-trade outcome. Opposite examples where vote trading is Pareto superior to no-trade can easily be constructed too<sup>8</sup>, and the literature eventually ran dry with the tentative conclusion that no general statement on the desirability of vote trading can be made.

Our algorithm leads us to the same conclusion, but we reach some unexpected results in special cases. In particular, when the committee is faced with only two proposals (and thus, since each proposal can either pass or fail, four possible outcomes), then for any number of voters and any (separable) preferences, the outcome associated with all stable vote allocations must be unique, is always Pareto optimal, is the Condorcet winner, if a Condorcet winner exists, and must be preferred by the majority to the no-trade outcome if it differs from it. These results hold whether trade is restricted to pairs of voters or coalitions are allowed. They are surprising because it has always been understood that vote trades' ambiguous welfare properties are due to the externalities inherent in the exchanges. But externalities are

<sup>6</sup>Tullock, 1959 and 1961, Downs, 1957, 1961.

<sup>7</sup>See also Coleman (1966), Haeferle (1970), Tullock (1970), and Wilson (1969).

<sup>8</sup>For example, Schwartz (1975).

clearly present in the two-proposal case, and yet the algorithm delivers an outcome with desirable welfare properties.

Approaching vote trading through a mechanical algorithm allowed us to make some progress by avoiding the difficulties of a strategic model. We have chosen this direction, however, for a second reason too: we conjectured that it may have predictive power. In the second part of the paper, we test the Pivot algorithm in the laboratory. The barter nature of the task, and thus the lack of a common unit of exchange, the changing profitability of trades in response to others' trades, the role of pivotality, all make the experiment unusually complex.<sup>9</sup> For this reason, we limit trades in the laboratory to pair-wise trades.

We study three treatments, corresponding to three sets of cardinal values for each voter over each proposal. All treatments have five voters, but differ in the number of proposals (two in treatment *AB*, and three in treatments *ABC1* and *ABC2*), and in the prediction of the Pivot algorithm. The Condorcet winner exists in all three cases; it coincides with the unique stable outcome reachable through the Pivot algorithm in treatments *AB* and *ABC2*; it differs in treatment *ABC1*.

Our experiment produces three main results. First, we find that stability is a useful predictive tool. In all treatments, two thirds or more of the vote allocations reached by experimental subjects are stable.

Second, although the final vote allocations provide some qualified support for the Pivot algorithm's ability to predict where vote trading will stop. Across all treatments, across all voters, across all proposals, in every single case in which the stable allocation reachable via the Pivot algorithm reflects a net purchase of votes, or a net sale, we observe it in the data. Yet, the final outcomes differ from the Pivot predictions, typically displaying more inertia around the no-trade outcome than the algorithm would predict.

The reason, and this is our third result, is that, contrary to the Pivot-algorithm, a large fraction of trades do not improve the myopic outcome for the traders but increase the number of votes held on high-value issues—what we define as “score-increasing trades”. Trades that increase the number of votes held by the winning side make further Pivot trades impossible and create a bias towards pre-trade out-

---

<sup>9</sup>To our knowledge, barter experiments are rare. Ledyard, Porter and Rangel (1994) is an example that demonstrates the challenges to both design and data analysis.

comes. Interestingly, score-increasing trades are not predictive of final vote allocations: trading stops when no opportunity for payoff gains remains, as in our stability concept, even when it is still possible to increase votes on high value proposals.

The first experimental paper studying vote trading was McKelvey and Ordeshook (1980). That paper reports results from a large series of experiments, all done face-to-face and under various protocols designed to allow either pair-wise only or coalitional trades, and either binding or non-binding agreements. The methodologies are different enough to make a direct comparison of results nearly impossible, and McKelvey and Ordeshook's focus on alternative cooperative solution concepts has no counterpart in our experiment.<sup>10</sup> Closer to our computerized experimental protocols are recent experiments on decentralized matching, in particular Echenique and Yariv (2013).<sup>11</sup> In that work, as in ours, a central finding is the extent to which the experimental subjects succeed in reaching a far-from-apparent set of stable matches. The set-up however differs substantially from ours, even within the perspective of matching theory: a two-sided matching problem with no externalities and fixed preferences, in Echenique and Yariv contrasting with the one-sided matching problem with externalities and evolving preferences in our case. In addition, the substantive questions we ask are specific to vote-trading, not matching.

The paper proceeds as follows. The next section presents our model and derives the model's theoretical predictions; section 3 discusses the experimental design; section 4 reports the experimental results, and section 5 concludes. Longer proofs are collected in the Appendix 1, and the instructions from a representative experimental session are in Appendix 2.

<sup>10</sup>Fischbacher and Schudy (2010) conduct a voting experiment to examine the possible behavioral role of reciprocity when a sequence of proposals come up for vote. There is no explicit vote trading, but voters can voluntarily vote against their short term interest on an early proposal in hopes that such favors will be reciprocated by other voters in later votes.

<sup>11</sup>Other related works are Nalbantian and Schotter (1995), Niederle and Roth (2011) and Pais, Pinter and Veszteg (2011). These papers have incomplete information and study the effects of different offer protocols and other frictions. Kagel and Roth (2000) study forces leading to the unraveling of decentralized matching.



## 2 The Model

Consider a committee of  $N$  (odd) voters who must approve or reject each of  $K$  independent binary proposals. The set of proposals is denoted  $P = \{1, \dots, k, \dots, K\}$ . Committee members have separable preferences summarized by a set of cardinal values  $Z$ , where  $z_i^k$  is the value attached by member  $i$  to the approval of proposal  $k$ , or the utility  $i$  experiences if  $k$  passes. Value  $z_i^k$  is positive if  $i$  is in favor of  $k$  and negative if  $i$  is opposed. Proposals are voted upon one-by-one, and each proposal  $k$  is decided through simple majority voting.

Before voting takes place, committee members can trade votes. The initial allocation of votes assigns one vote to each committee member over each proposal. We treat votes as if they were physical ballots, specialized by proposal—for example, imagine ballots of different colors for different proposals. A vote *trade* is thus modeled as an actual exchange of ballots, with no enforcement or credibility problem, where by exchange we mean that each trader must give away and receive at least one vote. After trading, a voter may own zero votes over some proposals and several over others, but cannot hold negative votes. We call  $v_i^k$  the votes held by voter  $i$  over proposal  $k$ ,  $V_i = \{v_i^k, k = 1, \dots, K\}$  the set of votes held by  $i$  over all proposals, and  $V = \{V_i, i = 1, \dots, N\}$  the profile (or *allocation*) of vote holdings over all voters and proposals.  $\mathcal{V}$  denotes the set of feasible vote allocations:  $V \in \mathcal{V} \iff \sum_i v_i^k = N$  for all  $k$  and  $v_i^k \geq 0$  for all  $v_i^k \in V$ .<sup>12</sup> The initial allocation of votes is denoted by  $V_0$ .

Given a feasible vote allocation  $V$ , we assume that at the time of voting, voters who attach positive value to a proposal cast all votes they own over that proposal in its favor, and voters who oppose it cast all available votes against it. We indicate by  $\mathbf{P}(V) \in P$  the set of proposals that receive at least  $(N+1)/2$  favorable votes, and therefore pass. We call  $\mathbf{P}(V)$  the *outcome* of the vote if voting occurs at allocation  $V$ . Note that with  $K$  independent binary proposals, there are  $2^K$  potential outcomes (all possible combinations of passing and failing for each proposal). Finally, we define  $u_i(V)$  as the utility of voter  $i$  if voting occurs at  $V$ :  $u_i(V) \equiv \sum_{k \in \mathbf{P}(V)} z_i^k$ .

Our focus is on the existence and properties of vote allocations that hold no

<sup>12</sup>Note that  $\sum_k v_i^k \neq K$  is feasible because we are allowing a voter to trade votes on multiple issues in exchange for one or more votes on a single issue. Of course, the aggregate constraint  $\sum_i \sum_k v_i^k = NK$  must hold.

incentives for further trading. We can then define:

**Definition 1** A pair of voters  $i, i'$  is said to **block**  $V$  if there exists a feasible vote allocation  $\hat{V} \in \mathcal{V}$  such that  $\hat{V}_j = V_j$  for all  $j \neq i, i'$ , and  $u_i(\hat{V}) > u_i(V)$ ,  $u_{i'}(\hat{V}) > u_{i'}(V)$ .

**Definition 2** An allocation  $V \in \mathcal{V}$  is **pair-wise stable** if there exists no pair of voters  $i, i'$  who can block  $V$ .

We can show immediately that a feasible allocation of votes that yields dictator power to a single voter  $i$  is trivially pair-wise stable: no exchange of votes involving voter  $i$  can make  $i$  strictly better-off; and no exchange of votes that does not involve voter  $i$  can make anyone else strictly better-off.<sup>13</sup> Hence:

**Proposition 1** A pair-wise stable vote allocation  $V$  exists for all  $Z, N$ , and  $K$ .

## 2.1 Dynamic adjustment.

Pair-wise stable allocations exist, but are they reachable through sequential decentralized trades? To answer the question, we need to specify the dynamic process through which bilateral trades take place. Our focus is on simple myopic algorithms.

We begin with the following definition:

**Definition 3** A trade is **minimal** if it consists of a minimal package of votes such that both members of the pair strictly gain from the trade.

Recall that a trade involves the exchange of at least one vote on each side.<sup>14</sup> Concentrating on minimal trades allows to "unbundle" complex trades into elementary trades. For any individual voter, multiple welfare-improving trades cannot be bundled, and zero-utility trades cannot be bundled with strictly welfare-improving trades.

<sup>13</sup>Other examples are easy to construct. For example, any allocation such that for all  $k$ ,  $q_{i_k}^k \geq \frac{n+1}{2}$ , where  $i_k$  is the voter who, among all, attributes highest value to winning on proposal  $k$ , is pair-wise stable.

<sup>14</sup>Thus surrendering votes to an ally with no myopic utility change is not a trade.

Although the literature does not make explicit reference to an algorithm, the sequential myopic trades envisioned by Riker and Brams (1973) and Ferejohn (1974) lend themselves naturally to such a formalization. In line with these earlier analyses, we define the *Pivot Algorithms* as sequences of pair-wise trades yielding myopic strict gains to both traders:

**Definition 4** A *Pivot Algorithm* is any mechanism generating a sequence of trades in the following way: Start from any vote allocation  $V_0$ . If there is no minimal pairwise (strictly) improving trade, stop. If there is one such trade, execute it. If there are multiple pairwise improving trades, choose one according to a possibly stochastic choice rule  $R$ . Continue in this fashion until no further improving trade exists

The definition above defines a whole family of algorithms, depending on the choice rule that is applied when there are multiple improving trades. Rule  $R$  specifies how the algorithm selects among multiple possible trades; for example,  $R$  may select each potential trade with equal probability (fully random); or give priority to trades with higher total gains; or to trades involving specific voters. The family of Pivot algorithms corresponds to the class of possible  $R$  rules, and individual algorithms differ in the specification of rule  $R$ . At this stage, it is not necessary to be more specific about  $R$ .

Pivot trades are not restricted to two proposals only: a voter can trade his vote, or votes, on one issue in exchange for other voters' vote(s) on more than one issue. The only constraint is the requirement that trades be minimal: zero-utility trades cannot be bundled with welfare improving trades. If a trade is welfare improving and minimal, it is a legitimate trade under Pivot.

A crucial property was anticipated by Riker and Brams and gives the name to our algorithm:

**Lemma 1** (Riker and Brams) Under the Pivot algorithms, all votes transferred must be pivotal.

**Proof.** Immediate from the requirement of minimal trades and the definition of Pivot algorithms.

## 2.2 Existence of stable vote allocations

The question we want to ask is whether a stable vote allocation is reachable through the Pivot algorithms. We define:

**Definition 5** An allocation of votes  $V$  is **Pivot-stable** and is denoted by  $V_T(R)$  if it is stable and reachable through a Pivot algorithm in a finite number of steps, following rule  $R$ .

Does a Pivot-stable allocation always exist? Surprisingly, the answer is clear-cut and positive. Pivot-stable vote allocations always exist, for the entire class of Pivot algorithms, independently of the rules  $R$  through which competing claims to trade are resolved. We can state:

**Theorem 1** For all  $K, N, Z$ , a Pivot-stable allocation of votes exists for all  $R$ .

**Proof.** Consider trades dictated by the Pivot algorithm. By Lemma 1, if a trade occurs at  $V_0$  it can only concern proposals that at  $V_0$  are decided by minimal majority. But by minimality of trade, it then follows that the same proposals must still be decided by minimal majority in any subsequent votes allocation  $V_t$ , with  $t > 0$ . But since  $V_0 = \{\mathbf{1}, \mathbf{1}, \mathbf{1}, \dots\}$ , no more than one vote is ever traded on any given proposal (although trades could involve bundles of proposals). Now consider voter  $i$  with values  $Z_i$  and absolute values  $|Z_i| \equiv X_i$ . We call  $i$ 's score at step  $t$  the function  $\sigma_{it}(X_i, V_{it})$  defined by:

$$\sigma_{it} \equiv \sum_{k=1}^K x_i^k v_{it}^k$$

where  $x_i^k$  is the (absolute) value  $i$  attaches to each proposal  $k$ , and  $v_{it}^k$  is the number of votes  $i$  holds on that proposal at  $t$ . If  $i$  does not trade at  $t$ , then  $\sigma_{it+1} = \sigma_{it}$ . If  $i$  does trade, then, by the argument above,  $i$ 's vote allocation must fall by one vote on some proposals  $\{k, k', \dots\}$  that  $i$  was winning and increase by one vote on some other proposals  $\{\tilde{k}, \tilde{k}', \dots\}$  that  $i$  was losing. Call the first set of proposals  $\mathbf{P}_{i,t}$  and the second  $\mathbf{P}_{-i,t}$ . Note that although the two sets may have different cardinality, by

definition of pair-wise improving trade,  $\sum_{k \in \mathbf{P}_{i,t}} x_i^k \leq \sum_{k \in \mathbf{P}_{-i,t}} x_i^k$  and, since a single vote is traded on each proposal,  $\sum_{k \in \mathbf{P}_{i,t}} x_i^k u_{it}^k \leq \sum_{k \in \mathbf{P}_{-i,t}} x_i^k u_{it+1}^k$ . Hence if  $i$  trades at  $t$ ,  $\sigma_{it+1} > \sigma_{it}$ : for all  $i$ ,  $\sigma_{it}(X_i, V_{it})$  must be non-decreasing in  $t$ . At any  $t$ , either there is no trade and the Pivot-stable allocation has been reached, or there is trade, and thus there are two voters  $i$  and  $i'$  for which  $\sigma_{it+1} > \sigma_{it}$  and  $\sigma_{i't+1} > \sigma_{i't}$ . But  $\sigma_{it}(X_i, V_{it})$  is bounded above and the number of voters is finite. Hence trade must stop in finite steps: a Pivot-stable allocation of votes always exists.<sup>15</sup> Note that we have made no assumptions on  $R$ , the rule through which trades are selected when multiple are possible. A Pivot-stable allocation of votes exists for any  $R$ .  $\square$

The generality of the result is surprising: a Pivot-stable allocation always exists, regardless of the number of voters and proposals, for all (separable) preferences, and regardless of the order in which different possible trades are chosen. As we said, the parallel to the matching literature is imperfect, and indeed no such result can be found there. In one-sided matching problems, it is well-known that a stable match may not exist.<sup>16</sup> When it does exist, it is not the case that any sequence of decentralized myopic matchings will converge to a stable matching. Cycles are possible. If preferences are strict, one converging sequence of matchings always exists, but if matchings are decentralized, guaranteeing convergence requires some randomness in the selection of blocking pairs: a random rule assigning a positive probability of selection to any blocking pair.<sup>17</sup> The difficulty of achieving stability is increased by the presence of externalities. We are not aware of comparable results for one-sided matching problems with externalities. In two-sided matching problems, guaranteeing the existence of a stable match in the presence of externalities requires a very stringent definition of blocking.<sup>18</sup> In our problem, the score function we have defined above is not subject to cycles. Because it is always non-decreasing in  $t$ , convergence to a stable allocation of votes is guaranteed for any selection rule

<sup>15</sup>It is not difficult to find the upper boundary on the number of trades needed to reach a Pivot stable allocation. It equals the maximum number of trades that could shift all individuals' votes to their respective highest-value proposal, or  $\left\lceil \frac{K(K-1)}{2} \right\rceil \left\lceil \frac{(N-1)}{2} \right\rceil$

<sup>16</sup>Gale and Shapley (1962).

<sup>17</sup>Diamantoudi et al. (2004). The result that randomness in the selection of the blocking pair induces convergence builds on Roth and Vande Vate (1990), who established it in the case of two-sided matching

<sup>18</sup>Sasaki and Toda (1996). Two individuals block an existing match if they strictly gain from matching with one another under any possible rematching by all others

among blocking pairs.

## 2.3 Preferences over stable outcomes

**Definition 6** An outcome  $\mathbf{P}(V)$  is **Pivot-stable** if it is achieved from a Pivot-stable allocation of votes

We denote by  $\mathcal{P}(V_T(R))$  the set of all stable outcomes reachable with positive probability through a Pivot algorithm with rule  $R$ . What are the welfare properties of  $\mathcal{P}(V_T(R))$ ? We have modeled vote trading through an algorithm, and our institution-free approach demands a welfare evaluation that is equally institution-free. We ask whether outcomes in  $\mathcal{P}(V_T(R))$  must belong to the Pareto set; whether they must include the Condorcet winner, if one exists, and more generally whether they can be ranked, in terms of majority preferences, relative to the no-trade outcome.

Our first set of answers is unexpectedly positive. Because we characterize results that hold for all  $R$ , we can use the simpler notation  $\mathcal{P}(V_T)$  with element  $\mathbf{P}(V_T)$ . We can show:

**Proposition 2** If  $K = 2$ , then, for all  $N$ ,  $Z$ , and  $R$ : (1)  $\mathbf{P}(V_T)$  is unique.<sup>19</sup> (2)  $\mathbf{P}(V_T)$  is Pareto optimal. (3) If a Condorcet winner exists, then  $\mathbf{P}(V_T)$  is the Condorcet winner. (4)  $\mathbf{P}(V_T)$  can never be the Condorcet loser. (5) If  $\mathbf{P}(V_T) \neq \mathbf{P}(V_0)$ , then a majority prefers  $\mathbf{P}(V_T)$  to  $\mathbf{P}(V_0)$ .

**Proof:** See Appendix I.

Proposition 1 is interesting because it highlights that the lack of Pareto optimality in vote trading examples, in particular Riker and Brams' paradox of vote trading, is not an immediate result of voting externalities. Externalities are not eliminated when  $K = 2$ , and yet the outcome of the Pivot algorithm (the same myopic vote-trading rule studied by Riker and Brams) is always Pareto optimal. Similarly, when  $K = 2$  vote trading performs well in terms of majority preferences.<sup>20</sup>

<sup>19</sup>Note that uniqueness of  $\mathbf{P}(V_T)$  does not imply uniqueness of  $V_T$ .

<sup>20</sup>Possibly, but not necessarily also in terms of total utilitarian welfare. In a finite electorate, results on utilitarian welfare depend on the distributions from which values are drawn.

In fact, there is another scenario in which, for all values  $Z$ , the Pivot-stable outcome is related to majority preferences:

**Proposition 3** *If  $N = 3$ , then for all  $K$ ,  $Z$ , and  $R$ : (1) If a Condorcet winner exists,  $\mathbf{P}(V_T)$  is unique and is the Condorcet winner. (2)  $\mathbf{P}(V_T)$  can never be the Condorcet loser*

**Proof:** See Appendix I

The intuition behind Proposition 3 is straightforward: with three voters, any Pivot trade between a pair reflects the majority's preferences. We know from Park (1967) and Kadane's (1972) that if there is a Condorcet winner, it can only be the no-trade outcome<sup>21</sup>, and thus with three voters there cannot be any trade under the Pivot algorithm. If the Condorcet loser exists, it must be such that all proposals are decided in the direction favored by the minority, but since any trade reflects the majority's preferences, such an outcome is impossible to reach.

The results from Pivot trading are less predictable in the general case:

**Proposition 4** *If  $K > 2$  and  $N > 3$ , then: (1) There exist  $Z$  such that for any  $R$  no outcome in the Pareto set is Pivot-stable. (2) There exist  $Z$  such that the Condorcet winner exists, but for any  $R$  it is not Pivot-stable*

**Proof:** We prove the two statements by example, and because the examples are simple and instructive, we report them here in detail. (1) Consider the following example, with  $K = 5$  and  $N = 5$ :

Each row in Table 1 is a proposal ( $A, B, C, D$ , and  $E$ ) and each column a voter (1, 2, 3, 4, and 5). Each cell  $\{k, i\}$  reports  $z_i^k$ , the value attached by voter  $i$  to proposal  $k$  passing. No voter is pivotal, and thus  $V_0$  cannot be blocked. The unique stable outcome is  $\mathbf{P}(V_T) = \mathbf{P}(V_0) = \{\emptyset\}$ , all proposals fail. Yet, all proposals failing is not Pareto optimal: it is Pareto-dominated by all proposals passing. The example suggests the importance of allowing for trades among coalitions of more than two voters, a point to which we return below.

(2) Consider the following example, with  $K = 3$  and  $N = 5$ :

<sup>21</sup>Because the majority must prefer the no-trade outcome to any outcome that differs from no-trade in the resolution of a single issue.

	1	2	3	4	5
A	10	-1	-1	-1	-1
B	-1	10	-1	-1	-1
C	-1	-1	10	-1	-1
D	-1	-1	-1	10	-1
E	-1	-1	-1	-1	10

Table 1: *Preference profile such that no outcome in the Pareto Set is Pivot Stable.*

	1	2	3	4	5
A	4	-7	1	-1	4
B	1	1	-4	4	-1
C	-3	4	2	-2	2

Table 2: *Preference profile such that the Condorcet Winner is not Pivot Stable.*

For the preference profile in Table 2,  $\mathbf{P}(V_0) = \{ABC\}$  is the Condorcet winner but this example has a unique Pivot-stable outcome  $\mathbf{P}(V_T) = \{A\}$ . It is not difficult to verify that there are three possible trade chains but all stop at  $\mathbf{P}(V_T) = \{A\}$ . Indicating in order the voters engaged in the trade, the proposals on which they trade votes (in lower-case letters)<sup>22</sup>, and, in parenthesis, the outcome corresponding to that allocation of votes, we can describe the three chains as  $\{\{13cb(A), 45bc(ABC), 23ab(C), 45ca(A)\}, \{23ab(C), 45ba(ABC), 13cb(A)\}, \text{ and } \{23ab(C), 45ca(A), 45bc(ABC), 13cb(A)\}\}$ .  $\square$

Note that result (2) in Proposition 3 immediately implies:

**Corollary 1** *If  $K > 2$  and  $N > 3$ , then there exist  $Z$  such that for all  $R$ ,  $\mathbf{P}(V_T) \neq \mathbf{P}(V_0)$ , but  $\mathbf{P}(V_0)$  is majority preferred to  $\mathbf{P}(V_T)$ .*

The first example is an immediate implication of Lemma 1: under the Pivot algorithm, trade can occur only between pivotal voters. If the vote allocation does not correspond to minimal majority, no pivotal voters exist. Thus the status quo is Pivot-stable, and delivers the unique Pivot-stable outcome; if such an outcome is Pareto-inferior, then the stable outcome does not belong to the Pareto set.

<sup>22</sup>For example, 13cb indicates that voter 1 acquires a C vote from voter 3, in exchange for a B vote.



The second example is more unexpected. Why does the positive result with  $K = 2$  not extend to a larger number of issues? Intuitively, the problem is that previous trades over some issues  $k$  and  $k'$  can make it impossible for a pair of voters to execute a different, desired trade over  $k'$  and  $k''$ . Thus, contrary to the  $K = 2$  case, the Pivot algorithm does not allow voters to exploit all opportunities for mutual agreements.

## 2.4 Coalitional Trades

A natural aspect of vote trading is the possibility of forming coalitions, indeed the incentive to do so. The experiment we describe below focuses on pair-wise trades, but our approach can be extended to the study of coalitions and sheds some light on the debates in the early literature on vote-trading. In this subsection, we derive two main results. First, the stability highlighted by our theorem on pair-wise trades does not generalize to vote-trading within larger coalitions of voters. Second, in our model, there is no logical connection between stability under coalitional trade and existence of the Condorcet winner.

We begin by redefining stability in the presence of coalitions. A coalition of voters  $C = \{i, i', i'', \dots\}$  is said to block  $V$  if there exists a feasible vote allocation  $\tilde{V} \in \mathcal{V}$  such that  $\tilde{V}_j = V_j$  for all  $j \notin C$ , and  $u_i(\tilde{V}) > u_i(V)$  for all  $i \in C$ . The allocation  $V \in \mathcal{V}$  is *coalition-stable* if there exists no coalition of voters  $C$  who can block  $V$ . Up to now, we have restricted  $C$  to be of size 2; here we allow  $C$  to have any size between 2 and  $N$ .

As in the case of pair-wise trades, the first observation is that a coalition-stable allocation always exists: a feasible allocation of votes that gives decision power to a single voter over all proposals remains trivially stable because no coalition that excludes the dictator can change the outcome, and the dictator cannot strictly gain from participating in any coalition. The interesting question is not whether a stable allocation exists, but rather whether it can be reached through the relevant extensions of our algorithm to coalitional trades.

To extend the algorithm, the previous definitions need to be amended. Keeping in mind that a vote trade must, by definition, include at least two voters and at least two issues, we define:

**Definition 7** A coalition-improving trade is **minimal** if it concerns: (1) the minimal package of votes such that all members of the coalition strictly gain from the trade; and (2) the minimal number of members such that the outcome corresponding to  $\hat{V}$  can be achieved.

**Definition 8** A **C-Pivot Algorithm** is any mechanism generating a sequence of trades in the following way: Start from any vote allocation  $V_0$ . If there is no minimal coalitional (strictly) improving trade, stop. If there is one such trade, execute it. If there are multiple coalitional improving trades, choose one according to a possibly stochastic choice rule  $R_C$ . Continue in this fashion until no further coalitional improving trade exists.

The *C-Pivot Algorithm* is the natural generalization of the Pivot Algorithm to coalitions. Note again that coalitions can be of any size and we have imposed no rule selecting among them in the order of trades, when several coalition-improving trades are possible. We call *C-Pivot stable* an allocation of votes that cannot be blocked by any coalition and is reachable via the C-Pivot algorithm in a finite number of steps, and define a *C-Pivot stable outcome* as an outcome  $\mathbf{P}^C(V_T)$  that corresponds to a C-Pivot stable allocation of votes.

In the presence of coalitions, stability becomes a more elusive goal:

**Proposition 5** There exist  $K$ ,  $N$ ,  $Z$ , and  $R_C$  such that the C-Pivot algorithm never converges to a stable vote allocation.

**Proof:** See Appendix I.

In the presence of coalitions, stability may fail because trades can be profitable for the coalition even when the pair-wise trades that are part of the overall exchange are not: coalition members benefits from the positive externalities that originate from the trades of other members. As a result, the score function defined in the proof of the Theorem in section 3 is no longer monotonically increasing in the number of trades, and the logic of that proof does not extend to coalition trades.

But if a C-Pivot stable allocation exists, does it have desirable welfare properties? A first, positive result is immediate:

**Proposition 6** *If a C-Pivot stable outcome exists, then it cannot be Pareto dominated.*

Regardless of the history of previous trades, if an outcome is Pareto dominated, then the coalition of the whole can always reach the Pareto superior outcome. But then the allocation corresponding to the Pareto dominated outcomes cannot be C-Pivot stable.<sup>23</sup>

But further results are more ambiguous:

**Proposition 7** *Consider  $Z$  such that the Condorcet winner exists. (1) If either  $K = 2$  or  $N = 3$ , then the C-Pivot stable outcome always exists, is unique, and coincides with the Condorcet winner. (2) If  $K > 2$  and  $N > 3$ , if a C-Pivot stable outcome exists, it need not coincide with the Condorcet winner*

**Proof.** (1) See Appendix 1. (2) Consider Example 2, in the proof of Proposition 4. There are two C-stable outcomes:  $\mathbf{P}(V_T) = \{A\}$ , reached through pair-wise trades as described earlier, and  $\mathbf{P}(V_T) = \{A, B\}$ , if we allow  $|C| > 2$ . The second stable outcome is reached through the following trades: after voters 2 and 3 have traded votes on  $A$  and  $B$ , a coalition of voters 1, 4, and 5 is formed; 4 gives an  $A$  vote to 1; 5 gives a  $B$  vote to 4, and a  $C$  vote to 1. Note that the trade is minimal, and the resulting vote allocation cannot be blocked. Hence  $\mathbf{P}(V_T) = \{A, B\}$ . Neither outcome is the Condorcet winner.  $\square$

Proposition 7 establishes that a C-stable outcome, when it exists, has no necessary link to the Condorcet winner. In some cases, ( $K = 2$  or  $N = 3$ ), the two must coincide; in others, ( $K > 2$  and  $N > 3$ ), the existence of the Condorcet winner gives no information about the welfare properties of C-stable outcomes. The result is driven by two central assumptions of our model: (1) vote trades are binding, and (2) voting occurs proposal-by-proposal.

<sup>23</sup>Note that if the coalition trade is not minimal, it can be made so by eliminating redundant trades or traders.

### 3 The Experiment

The experiment was run at the Columbia Experimental Laboratory for the Social Sciences (CELSS) in November 2014, with Columbia University students recruited from the whole campus through the laboratory’s ORSEE site. No subject participated in more than one session. After entering the computer laboratory, the students were seated randomly in booths separated by partitions; the experimenter then read aloud the instructions, projected views of the computer screens during the experiment, and answered all questions publicly.<sup>24</sup>

Because the design of the trading platform presents some challenges, we describe it here in some detail.<sup>25</sup>

At the start of each treatment, a subject saw on his computer screen the matrix of values, denominated in experimental points, and the vote allocation. To help intuition, the two alternatives for each issue—Pass or Fail—were identified with two colors—Orange or Blue—, and each individual’s values were written in the color of the individual’s preferred alternative.<sup>26</sup> The screen also showed the votes totals and the points the subject would win if voting were held immediately. Each subject started with one vote on each issue.

After having observed the matrix of values and the current vote allocation, a subject could post a bid for a vote on one of the issues, in exchange for his vote on a different issue. The bid appeared on all committee members’ monitors, together with the ID of the subject posting the bid. A different subject could then accept the bid by clicking the offer and highlighting it.<sup>27</sup>

A central feature of vote trading is that the preferences and vote holdings of the specific individuals making a trade determine the effect of the trade. Contrary to standard market experiments, then, subjects must not only post potentially profitable bids, but also consider the specific identity of their trading partner. In adapting the bidding platform used in market experiments, we added a confirmation step.

<sup>24</sup>A copy of the instructions and sample trading screens are in an online Appendix.

<sup>25</sup>The computerized trading platform was implemented the Multistage program, an open source software developed at Caltech’s Social Science Experimental Laboratory (SSEL). The software is available for public download at <http://multistage.ssel.caltech.edu:8000/multistage/>

<sup>26</sup>Thus all experimental values were positive and indicated earnings from one’s preferred alternative winning (relative to zero earning if it lost).

<sup>27</sup>We reproduce in Appendix B three screenshots shown during instructions.

After a bid is accepted, a window appears on the bidder's screen detailing the effects of that specific trade—what the outcome would be upon immediate voting—and asking the bidder to confirm or reject the trade. If the trade is rejected, a message appears on the screen of the rejected trade partner, informing him of the rejection.

After a trade was concluded, the vote tally on each issue was updated and conveyed to all subjects via a specific message on all screens. The message also reported the post-trade voting outcome if voting were to occur immediately. Note that the value matrix and the updated vote holdings were always present on the screen.

The market was open for three minutes.<sup>28</sup> However, in a market where each concluded trade can trigger a new chain of desired trades, it is particularly important to ensure that all desired trades have the time to be executed. Thus the time limit was automatically extended by 10 seconds whenever a new trade occurred within 10 seconds of the expected closure.

Only trades of a single vote on one issue against a single vote on a different issue were allowed, again to limit the complexity of the task. No bid could be posted if a subject did not have enough votes to execute it if accepted; thus a voter could post multiple bids only as long as he had enough votes to execute them all, had all been accepted. Posted bids could be canceled at any time, an important feature in a market where somebody else's executed trade can make an existing posted bid suddenly unprofitable.

Once the market closed, voting took place automatically, with all votes on each issue cast by the computer in the direction preferred by each subject. Then a new round started.

The experiment consisted of three treatments, *AB*, *ABC1*, and *ABC2*, each corresponding to a different matrix of values. In all three treatments, the size of the voting committee was five ( $N = 5$ ), while the number of issues depended on the treatment:  $K = 2$  in treatment *AB*, and  $K = 3$  in treatments *ABC1*, and *ABC2*. In each committee, subjects were identified by ID's randomly assigned by the computer, and issues were denoted by *A* and *B* (in treatment *AB*), and *A*, *B* and *C* (in treatments *ABC1* and *ABC2*). Each session started with two practice rounds; then three rounds of treatment *AB*, and then five rounds each of *ABC1*

---

<sup>28</sup>Two minutes in treatment *AB*, with two issues only.

and  $ABC2$ , alternating the order.<sup>29</sup> We did not alternate the order of treatment  $AB$  because its smaller size ( $K = 2$ ) made it substantially easier for the subjects, and thus we used it as further practice before the more complex treatments. This is also the reason for the smaller number of rounds (three for  $AB$ , versus five for  $ABC1$  and  $ABC2$ ).

Committees were randomly formed, and ID's randomly assigned at the start of each new treatment, but the composition of each group and subjects' ID's were kept unchanged for all rounds of the same treatment, to help subjects learn. All but one sessions consisted of 15 subjects, divided into three committees of five subjects.<sup>30</sup> At the end of each session, subjects were paid their cumulative earnings from all rounds, converting experimental points into dollars via a preannounced exchange rate, plus a fixed show-up fee. Each session lasted about 90 minutes, and average earnings were \$34.

We designed the three treatments according to the following criteria. First, we wanted a  $K = 2$  treatment, both as further training for the subjects and because of the sharp theoretical predictions of the Pivot algorithm in this case. Second, we chose value matrices for which the stable outcome reachable via Pivot trades is unique but requires multiple trades. In  $AB$ , the path to stability is unique, while in both  $ABC1$  and  $ABC2$  the Pivot stable outcome can be reached via multiple paths, with no path being clearly focal. Third, we chose matrices such that not only is the Pivot stable outcome unique, but the stable vote allocation reached via Pivot trades is unique, even with multiple possible trading paths. Fourth, we designed matrices for which the Condorcet winner exists, but need not correspond to the Pivot stable outcome: it does in  $AB$  (by necessity—see Proposition 2), and in  $ABC2$  (by construction), but not in  $ABC1$ . Finally, we wanted  $ABC1$  and  $ABC2$  to be superficially very similar and to have Pivot trading paths of similar multiplicity and length, allowing us to test whether the different force of attraction of the Condorcet winner predicted by the theory is reflected in the data. Note that we do not specify  $R$ , the selection rule when multiple trades are possible, but let the experimental subjects select which trades to conclude. Our theoretical results hold for all  $R$ .

The three preference profiles used in the experiment are given in Table 3.

<sup>29</sup>Two of the sessions had only two treatments:  $AB$  and  $ABC1$  in one case, and  $A$  and  $ABC2$  in the other.

<sup>30</sup>One session had only ten subjects, divided into two groups.

AB					
	1	2	3	4	5
A	49	-29	-29	12	-12
B	12	-12	-49	29	49

ABC1					
	1	2	3	4	5
A	23	-23	10	-10	23
B	-10	-10	23	-23	10
C	18	-18	-18	18	-18

ABC2					
	1	2	3	4	5
A	-21	15	-9	21	9
B	15	9	15	-15	-15
C	-9	-21	21	9	21

Table 3: *Preference profiles used in experiment*

Session	Treatments	# Subjects	# Groups	# Rounds
s1	AB, ABC1, ABC2	10	2	3,5,5
s2	AB, ABC2, ABC1	15	3	3,5,5
s3	AB, ABC1, ABC2	15	3	3,5,5
s4	AB, ABC2, ABC1	15	3	3,5,5
s5	AB, ABC2	15	3	3,5
s6	AB, ABC1	15	3	3,5

Table 4: Experimental Design. Note: A programming error in sessions s5 and s6 made the last five rounds of data unusable.

In all three cases, the initial vote allocation  $V_0$  is unstable. In the case of matrix AB,  $\mathbf{P} = \{B\}$  is the Condorcet winner and the unique Pivot-stable outcome. The Pivot algorithm follows a unique path, of length two (i.e. consists of a sequence of two trades). Matrix ABC1 has identical properties to the matrix of values discussed in the proof of Proposition 4. The Condorcet winner exists and corresponds to  $\mathbf{P} = \{A\}$ , but the unique Pivot-stable outcome is  $\mathbf{P} = \{A, B, C\}$ . In matrix ABC2, the Condorcet winner is  $\mathbf{P} = \{A, B, C\}$ , and corresponds to the unique Pivot stable outcome. With both matrices ABC1 and ABC2, the Pivot algorithm can follow three different paths, and for both matrices two of these paths have length four, and one has length three.<sup>31</sup>

Table 4 reports the experimental design.

<sup>31</sup>The possible paths are detailed in the Appendix.

Treatment	Tot trades	groups $\times$ rounds	Mean trades	Median	s.d	Max	Pivot
<i>AB</i>	115	51	2.25	2	1.92	13	2
<i>ABC1</i>	211	70	3.0	3	1.67	9	3,3,4
<i>ABC2</i>	175	70	2.5	2	1.36	7	3,3,4

Table 5: Number of trades.

## 4 Experimental Results

### 4.1 Trading

How much trading did we see? Table 5 reports basic statistics on observed trades. "Pivot" refers to the predicted number of trades under the Pivot algorithm. The unit of analysis is the group per round.

A histogram of the number of trades per treatment (Figure 1) shows clearly the higher frequency of few trades in the *AB* treatment, with  $K = 2$ . Between the two  $K = 3$  treatments, *ABC2* has consistently higher fractions of low trades, but the differences are not striking—56 percent of rounds end with two or fewer trades in *ABC2*, as opposed to 41 percent in *ABC1*, and 80 percent end with three or fewer trades in *ABC2*, as opposed to 76 percent in *ABC1*. In all treatments, few rounds include five or more trades.

As expected, the bidder's option of rejecting trades, and thus discriminating over who accepted the original bid, was important. In columns 2-4 of Table 3, we report the total number of bids, how many of these found a taker in the market, and how many of these acceptances were then rejected by the bidder. A large fraction of all posted bids found a counterpart—from a minimum of 77 percent in *ABC2* to more than 95 percent in *AB*—but about a third of these accepted trades were rejected by the bidder—32 percent in *A*, 29 percent in *ABC1*, and 34 percent in *ABC2*. As the last column of the table shows, some rejected trades were associated with a strict increase in myopic payoff for the bidder, but the number is small—between 10 and 20 percent of rejections in all treatments.

Whether in terms of number of trades or of any other variable studied below, the data show no evidence of learning or of order effects—behavior appears very consistent across rounds, and regardless of whether *ABC1* or *ABC2* was played



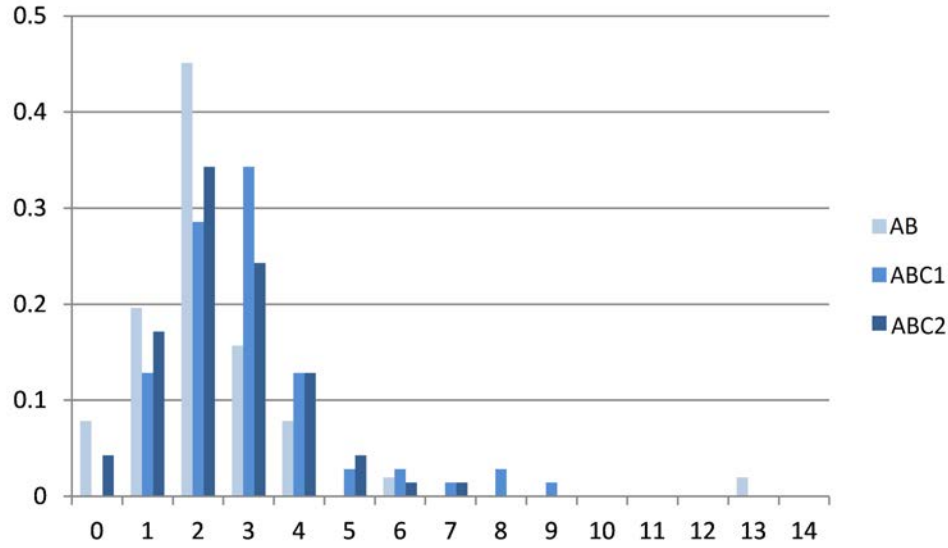


Figure 1: Number of trades. Frequencies

Treatment	Tot bids	Accepted	Rejected by bidder	Rejected with payoff gain
<i>AB</i>	177	169	54	6
<i>ABC1</i>	368	296	85	15
<i>ABC2</i>	345	267	92	11

Table 6: Bids, accepted bids, and rejected trades. Tot bids excludes canceled bids.

first. Thus we present the experimental results aggregating over rounds and order.

## 4.2 Stability

Our point of departure is the definition of stable vote allocations. Is the stability requirement satisfied in the vote allocation to which our subjects converge at the end of each round? Figure 2 shows the CDF of steps to stability for the three treatments, in blue, as well as in 5,000 simulations with random trading, in grey, with the dotted line. The horizontal axis measures the minimal number of Pivot trades necessary to reach stability, and the vertical axis the proportion of final vote allocations not further from stability than the corresponding number of trades.

The fraction of stable vote allocations in the experimental data was 76 percent in *AB*, and 64 percent in both treatments *ABC1* and *ABC2*. In all treatments, more

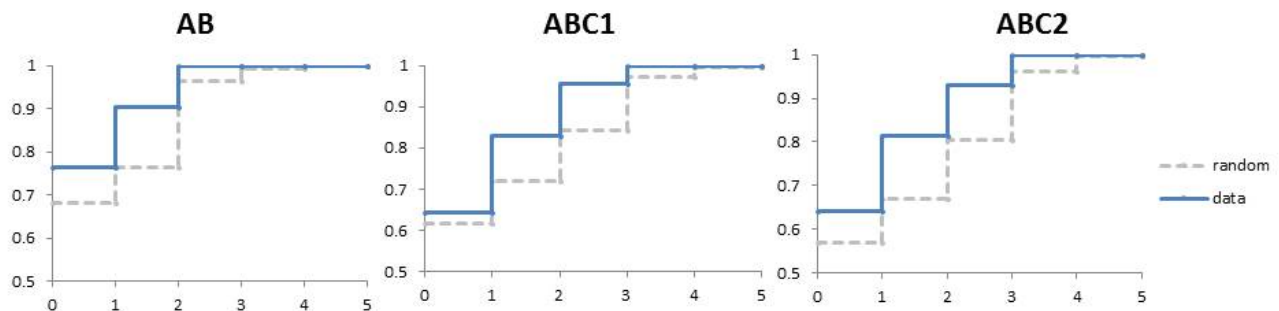


Figure 2: Steps to stability. Cumulative distribution functions.

than 80 percent of all vote allocations were within one step (one trade) of stability, although the figure also shows the predictably easier convergence to stability in the *AB* treatment, with only two proposals. In all three treatments, the distribution corresponding to random trading FOSD's the distribution for the experimental data.

The simulation of random trades provides the yardstick of comparison for our data. We will use it repeatedly in what follows, and it is worth describing the methodology in some detail. In each treatment, we constructed the random trades by randomly selecting an individual, one or two issues (in the two- and three-issue treatments, respectively), a partner, and a direction of trade, all with equal probability, and enacting the trade as long as both traders' budget constraints are satisfied. If budget constraints are violated, we cancel the proposed trade and restart. In each group, a trade occurs with specified probability over a short time interval, with both parameters calculated to match the observed average length of rounds and the

average number of trades in the treatment.<sup>32</sup> For each treatment, we repeated the procedure 5,000 times, each time focusing on a group.<sup>33</sup>

Figure 2 reports information on the stability of the vote allocations reached at the end of trading. But our data also give us information on dynamic convergence. Do successive trades move the vote allocation towards stability?

Figures 3 and 4 show, for each treatment, the dynamic path of the vote allocation, as captured by the succession of trades. The horizontal axis measures time, in seconds. A marker corresponds to a trade. Thus, for any given marker, the horizontal axis indicates when the trade took place, within the maximal round length observed in the data for each treatment.<sup>34</sup> The vertical axis measures distance from stability, defined, as in Figure 2, by the minimal number of Pivot trades necessary to reach a stable allocation. Such number is calculated first for the vote allocation characterizing each group in the treatment at that moment in that round, and then averaging over the groups. The figure is drawn pooling over all groups and all sessions, for given treatment, and each curve, with its own shade and marker symbol, reports data from the same round (1-3 for *AB* and 1-5 for *ABC1* and *ABC2*). The jumps between dots are relatively small because a trade concerns a single group, while the others' vote allocations remain unchanged.

All curves decline, almost perfectly monotonically, showing the dynamic convergence towards stability. To help us evaluate such convergence, the black curve in each panel reports the steps from stability calculated from the 5,000 simulations with random trading.

After the first minute, in all three treatments, the curve corresponding to random trades remains higher than the curve corresponding to any round of experimental

<sup>32</sup>Given the average length of a round in the treatment, time is divided into a grid of 100 cells, and in each cell a group can trade with probability  $p$ , such that  $100p$  equals the mean number of trades per round in the treatment.

<sup>33</sup>Note that random trading is a demanding comparison when applied to the stability of vote allocations. The reason is that a large fraction of feasible trades take the vote allocation away from minimal majority, and hence make pivot trades impossible, and the allocation stable. For example, in treatment *AB*, where breaking minimal majority on a single issue is sufficient to induce stability, a single random vote trade from any unstable allocation has never less than a 30 percent chance of inducing stability.

<sup>34</sup>The trading period lasts 180 seconds, but there is a 10 second delay with each trade, to give time to subjects to study the new vote allocations. In addition, 10 more seconds are added at the end of the period if any trade takes place in the last 10 seconds.

data.<sup>35</sup> Notice also the lack of learning in the data—there is no systematic difference between earlier and later rounds.<sup>36</sup>

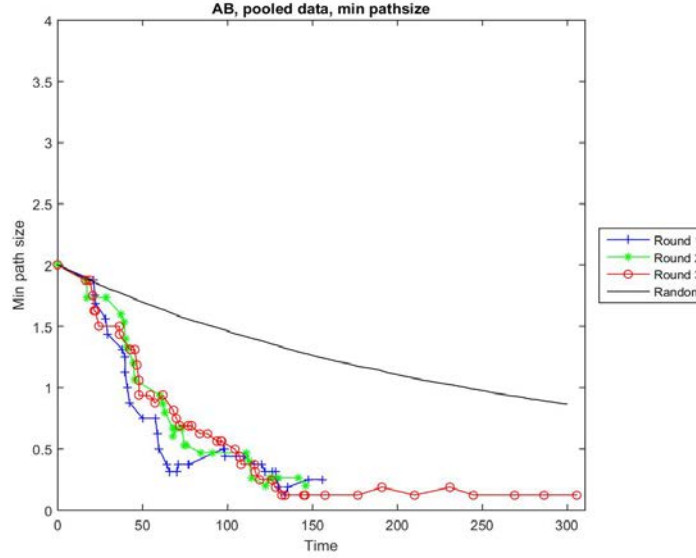


Figure 3: Dynamic convergence to pivot stable outcomes. Data vs. Random. AB matrix.

### 4.3 Vote Allocations

For all three value matrices used in our experiment, the Pivot algorithms predict a unique stable vote allocation. Is such an allocation reached by the experimental subjects? Figure 5 reports the number of votes held by each voter at the end of a round, averaged over all rounds of the same treatment. Each panel corresponds to a treatment and reports the number of votes by voter ID, i.e. by the vector of values corresponding to each column of the value matrix. The blue columns represent the experimental data, the grey columns the Pivot prediction, and the red line the no-trade status quo (or equivalently, the average vote holding after random trading).

<sup>35</sup>With the exception of two trades in round 5 in *ABC2*.

<sup>36</sup>To verify that results were not driven by averaging, we computed CDF's of steps to stability for the data and for the random simulations, as in Figure 1, at all 30-second intervals. In all treatments and at all times, the CDF corresponding to random trading FOSD's the CDF from the data.

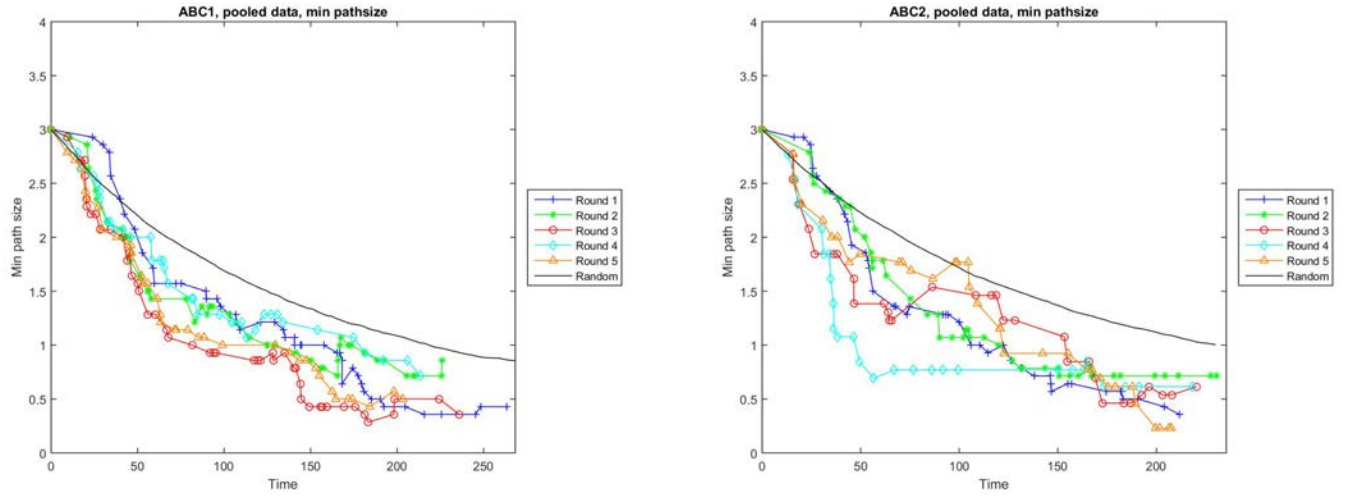


Figure 4: Dynamic convergence to pivot stable outcomes. Data vs. Random. ABC1, ABC2.

The figure reports data from all rounds, but remains effectively identical if we select stable vote allocations only.

It is clear from the figure that the vote distribution in the data is less sharply variable across issues than theory predicts, as we would expect in the presence of noise. Yet, the qualitative predictions are strongly supported. There are five voters in each treatment, holding votes over two (in *AB*) or three issues (in *ABC1* and *ABC2*)—a total of forty points. Of these forty, the theory predicts that 14 should be above 1—the voter should be a net buyer over that issue— and 15 below 1—the voter should be a net seller. The prediction is satisfied in *every* single case, across all treatments. When the theory predicts holding a single vote—11 cases for which the voter should exit trade with the same number of votes held at the start—, the data show three cases where the average vote holding is below 1, five where it is above, and three where it is effectively indistinguishable from 1. On average, our subjects hold 0.56 votes when the theory predicts 0; 1.05 when the theory predicts 1, and 1.43 when the theory predicts 2.<sup>37</sup>

<sup>37</sup>The theory predicts that voter 3 in treatment *ABC1* should hold three votes.

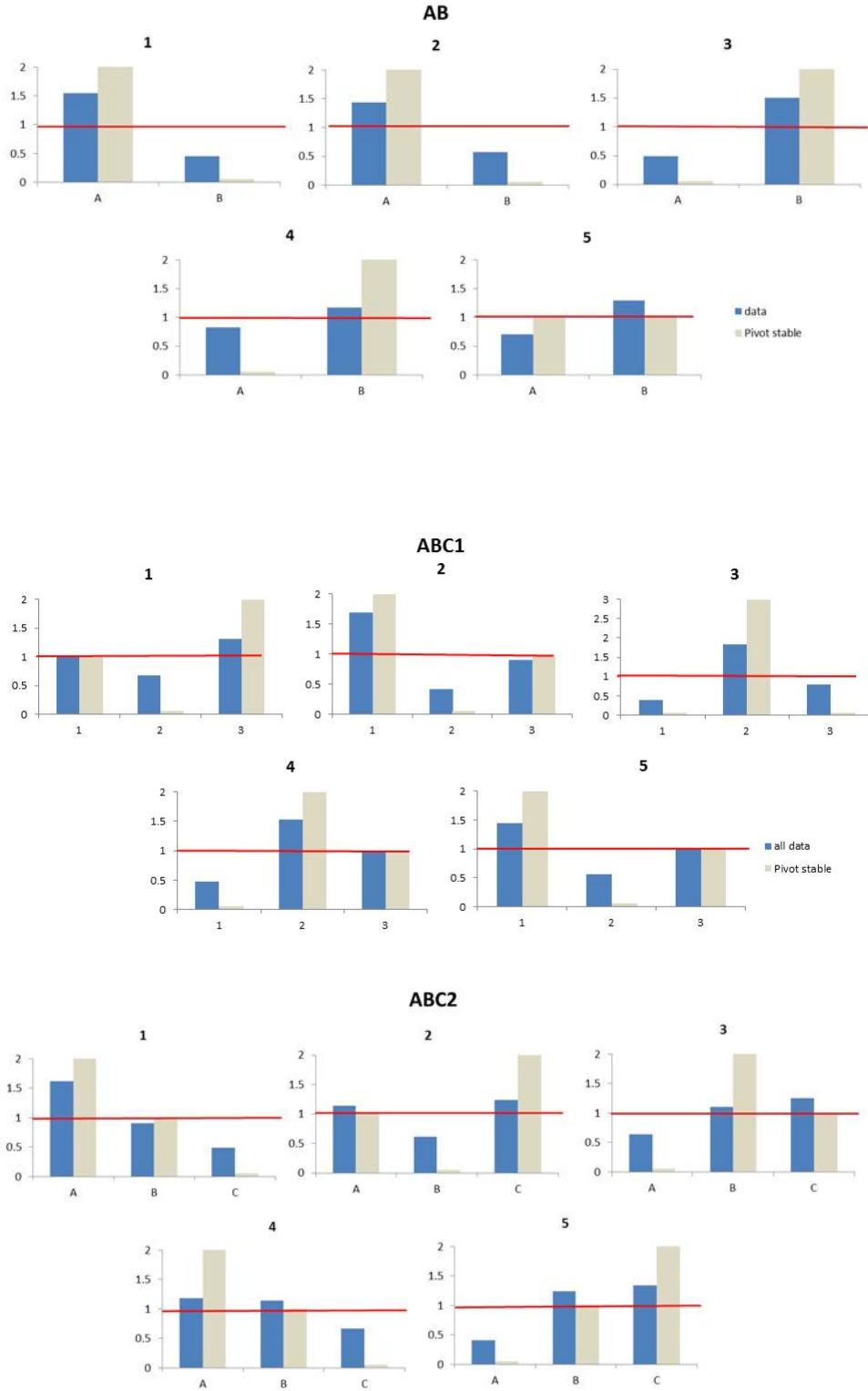


Figure 5: Average vote allocations at the end of each round, by voter type. Top Panel AB  
Middle Panel ABC1. Bottom Panel ABC2

## 4.4 Trades

According to our results so far, final vote allocations tend to be stable; dynamic trading moves towards stability, and final individual vote holdings mirror qualitatively the properties of Pivot-stable allocations. But can we say more about the specific trades we see in the lab? In particular, are these trades compatible with the Pivot algorithm?

### 4.4.1 Pivot trades.

The class of Pivot algorithms is a class of mechanical selection rules among feasible pair-wise trades. It is not a model of individual behavior. Accordingly, it should be tested not on individual trades, but on binary trades—i.e. by considering the fraction of all trades associated with myopic strict increases in payoff for both traders. We plot such a fraction in Figure 6. The blue columns correspond to the experimental data, the light grey columns to the simulations with random trading, and the error bars indicate 95 percent confidence intervals (under the null of random trading).<sup>38</sup>

The figure shows clearly the subjects' search for gains. With random trading, the frequency of payoff gains for both traders is 3 percent in  $AB$  and 1 percent in  $ABC1$  and  $ABC2$ , or less than one fifth of what we observe in  $AB$ , and less than one tenth in  $ABC1$  and  $ABC2$ . In all cases, the probability that the data are generated by random trades is negligible.

But if the trading behavior of the experimental subjects is not random, it is also true that the fraction of trades consistent with the Pivot algorithm is small: 17 percent in  $AB$ , 26 percent in  $ABC1$  and 18 percent in  $ABC2$ . Which other trades are subjects concluding?

### 4.4.2 Other trades.

We find that a much larger share of the data can be explained by extending the Pivot algorithm in one of two directions. First, while the Pivot algorithm selects trades with strict gains in payoffs, in every treatment more than 40 percent of all trades result in no change in payoff for either trader. Zero-gain trades are trades involving

<sup>38</sup>Note that under the null all observations are independent. Thus no correction for correlation is required.

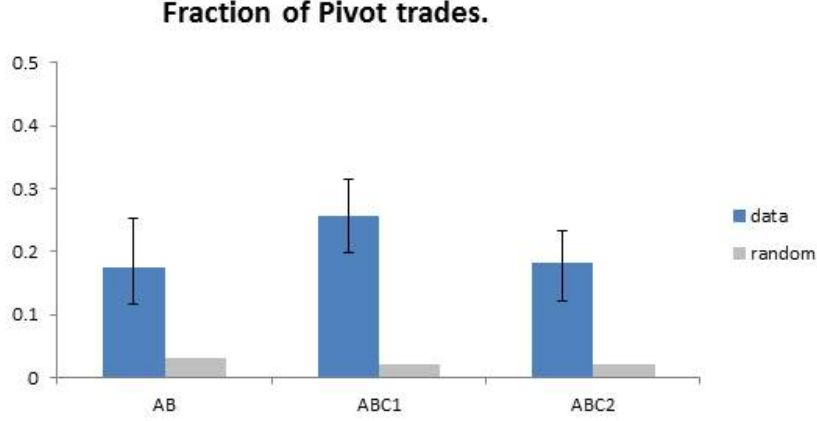


Figure 6: Fraction of Pivot Trades

non-pivotal votes, and thus preserving the status quo outcome; they could be the result of buying votes from allies with weak preferences, for example, or of buying losing votes, to strengthen one's favorite side's margin of victory. No (myopic) rationality requirement is violated by trades that are only weakly-improving, either for one or both traders, and our algorithm could be extended to accommodate such trades.

Second, as we note in the proof of the Theorem, every Pivot trade corresponds to an increase in the score function  $\sigma_{it}(X_i, V_{it})$  for the two traders involved.<sup>39</sup> But not all increases in score correspond to Pivot trades: trades that shift votes from low to high-value proposals do not cause strict myopic payoff gains if they do not change the resolution of the high-value proposals, either because they continue to be lost or

<sup>39</sup>Recall that  $\sigma_{it}(X_i, V_{it})$  is defined by:

$$\sigma_{it} \equiv \sum_{k=1}^K x_i^k v_{it}^k$$

where  $x_i^k$  is the absolute valuation attached by  $i$  to proposal  $k$  passing, and  $v_{it}^k$  is the number of votes on  $k$  held by  $i$  at  $t$ .



because they were already won. Such trades could reflect difficulties understanding pivotality, but could also mirror behavior that is more forward-looking than Pivot algorithms. Myopic gains are evaluated assuming voting occurred immediately. In fact, in the uncertain and complex environment of our experiment, subjects may want to accumulate votes on high value proposals, regardless of their resolution under immediate voting, because they conjecture that further trades are likely to take place before voting actually occurs.

Figure 7 shows, for each treatment, the fraction of binary trades consistent with Pivot trades (in dark blue), weak payoff increases for both traders (light blue), and score increases, again for both traders (in orange).<sup>40</sup>

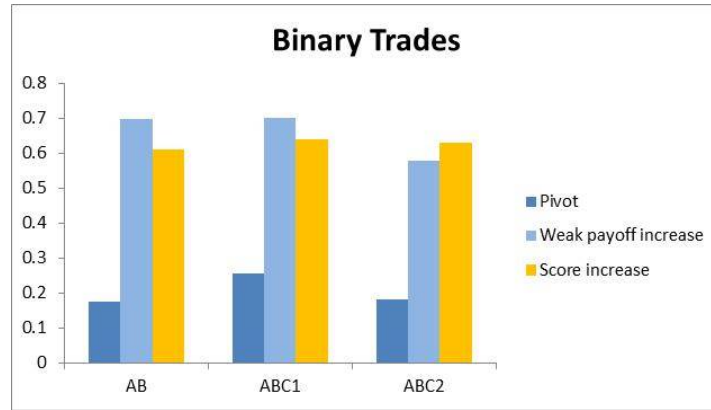


Figure 7: Types of Trades

By construction, Pivot trades are a subset of both of the other two categories, and thus must explain a lower fraction of observed trades. What is surprising is how much smaller. The figure shows that Pivot trades are of the order of one third of all weakly-payoff-improving trades in treatments *A* and *ABC2*, and about two fifths in treatment *ABC1*. Similar numbers apply to score-improving trades.

The frequency of different types of trades is informative, but what we need to understand is the intentionality of such trades. As we remarked about Figure 6, Pivot trades are not very frequent, but they appear intentional: they cannot be explained by random trading. Is that true of other types of trades?

Figure 8 plots, for the representative case of the *AB* treatment, the observed fractions of Pivot trades, zero-payoff change trades, and score-increasing-not-Pivot trades, together with the corresponding fractions under random trading and the 95

<sup>40</sup>The experimental matrices do not allow for weak score increases.

percent confidence interval under the null hypothesis of random trading.

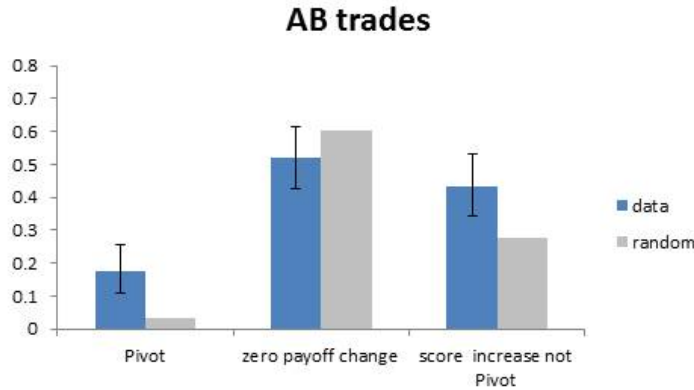


Figure 8: AB trades by type v/s random.

The figure makes clear that although the fraction of zero-payoff changing trades is large, we cannot rule out that it is the result of noisy trading: because all non-pivotal trades have zero effect on payoffs, for any given vote distribution a large share of feasible trades belongs to this class and thus is chosen under random trading. The figure does show, however, that this is not true for non-Pivot-score-increasing trades: the fraction observed in the data is significantly higher than under random trading ( $p < 0.0001$ ).

We can make these observations more precise through a simple statistical model.

#### 4.4.3 A simple statistical model

The model does not aim at explaining behavior but at classifying the types of trades, lending some rigor to the comments suggested by the figures. In line with the data just reported, we suppose that executed trades are selected according to four myopic criteria, synthetic summaries of the rules followed by the pairs of traders: (1) Pivot trades; (2) zero-payoff changing trades; (3) score-improving trades; (4) some other criterion we ignore, and such that the trade appears to us fully random. When

	$AB$		$ABC1$		$ABC2$	
$probP$	0.06	[0, 0.14]	0.19	[0.13, 0.25]	0.11	[0.05, 0.17]
$prob0$	0.11	[0, 0.23]	0.07	[0, 0.16]	0	[0, 0.10]
$probS$	0.41	[0.28, 0.55]	0.34	[0.25, 0.43]	0.39	[0.29, 0.49]
$probR$	0.42	[0.27, 0.57]	0.40	[0.29, 0.52]	0.50	[0.35, 0.59]

Table 7: *Model parameter estimates with 95% confidence intervals*

executing a trade, each pair of traders follows one of these rules. Each trade can then be written in terms of the probability of following the four criteria:  $probP$  for Pivot trading;  $prob0$  for zero-payoff changing trades,  $probS$  for score improving trades, and  $probR$  for random trades. Call  $T_t$  the set of all trades feasible at  $t$ , where a trade is defined by a pair of traders, a pair of proposals, and the direction of trade. Similarly, call  $T_t^P$  the set of all feasible Pivot trades,  $T_t^0$  the set of all feasible zero-payoff trades, and  $T_t^S$  the set of all feasible score-improving trades. Suppose that we observe a Pivot trade. The probability of such a trade equals  $probP/|T_t^P| + probS/|T_t^S| + probR/|T_t|$ . Similarly, the probability of a score-improving but not Pivot trade is given by  $probS/|T_t^S| + probR/|T_t|$ . Assuming that different trades are independent, the likelihood of observing the data set is simply the product of the probabilities of each trade. The probabilities  $probP$ ,  $prob0$ ,  $probS$ , and  $probR$  can then be estimated immediately through maximum likelihood. The only challenge is that the sets of feasible trades,  $T_t$ ,  $T_t^P$ ,  $T_t^0$ , and  $T_t^S$ , all evolve over time, as budget constraints become binding, and the changes in vote allocations alter the payoff effects of different vote exchanges.

We report our estimates in Table 7, together with the 95 percent confidence intervals.<sup>41</sup>

According to our statistical model, trade is very noisy and, as Figure 8 lead us to expect, there is no evidence of intentional zero-profit trades in any of the three treatments (in all treatments the 95 percent confidence interval for  $prob0$  includes 0). There is however a significant probability of Pivot trades in treatments  $ABC1$  and  $ABC2$ , and of score-improving trades in all three treatments. Again as implied by the figures,  $probP$  and  $probS$  are not fully collinear and can be estimated separately.

<sup>41</sup>We constructed the confidence intervals by bootstrapping the data and estimating the model's parameters 1000 times. The changing sets of different types of trades make the procedure computationally demanding

Score-improvement is a less powerful hypothesis than Pivot trading but explains a larger fraction of the data.

#### 4.4.4 Score-improving trades

Observing that a rising score can explain a substantial fraction of the experimental trades does not imply that the increase in score is the final objective pursued by our subjects. A first reason to be skeptical is the frequency of rejected trades reported in Section 4.1. Recall that about a third of all accepted bids are rejected by the bidder. Contrary to payoff changes, score increases do not depend on the identity of the trading partner: if score increases were the goal of the trades, they could be secured by the bidder and there would be no reason to reject any partner.<sup>42</sup>

A second cause for doubt comes from investigating whether subjects have indeed exploited all opportunities for score increases when trade comes to an end. We have defined stability as the absence of any feasible pairwise strictly payoff-improving trade. We can construct the similar concept of *score stability*, defined as the absence of any feasible pairwise score-improving trade, and enquire whether score stability is a useful characterization of final vote allocations.<sup>43</sup> Figure 9 plots the CDF's of minimal steps from score stability in the three treatments (in orange), together with the CDF of minimal steps to payoff stability (in blue).

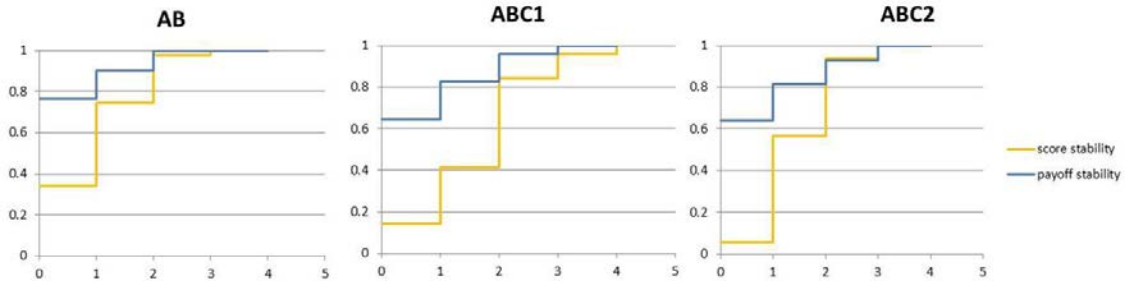


Figure 9: CDF's score and payoff stability

Score stability is a much weaker explanation of final vote allocations than payoff

<sup>42</sup>In fact, in all treatments more than two-thirds of the trades rejected by the bidder would have caused the bidder an increase in score.

<sup>43</sup>Note that a score-stable allocation always exists in pair-wise trading (by the proof of the Theorem).



different in  $ABC1$  and  $ABC2$ ). A star indicates the Condorcet winner, and a dot the Pivot stable outcome.

The figure shows two immediate regularities. First, in all treatments, the Condorcet winner is the most frequent outcome, whether we consider all outcomes, or stable outcomes only. Second, in all treatments, the frequency of outcomes correlates positively and significantly with aggregate payoffs. However, because both the Condorcet winner and aggregate payoffs also correlate perfectly with persistence of pre-trade outcomes<sup>44</sup>, both results may reflect the inertia built into the market by the frequent zero-gain trades.

In terms of Pivot predictions, we see a higher frequency of the Condorcet winner, relative to the second most frequent outcome, in treatments  $AB$  and  $ABC2$ , where the Condorcet winner is Pivot-stable. And among stable outcomes we see a small spike in the frequency of outcome  $\{A, B, C\}$  in treatment  $ABC1$  where it is Pivot-stable, relatively to the outcome's low payoff-rank. On the whole, however, the clean predictions on outcomes derived from the Pivot algorithm are not evident in the data. Contrary to a goods market, the outcomes of vote-trading are very sensitive to deviations—one subject's missed trading opportunity affects the final result of voting for all. As shown by Figure 11, the outcomes we observe are qualitatively in line with the trades' characteristics highlighted by the statistical model.

The figure reports the frequency of different outcomes in the data (considering here all final vote allocations, whether stable or unstable) and, in columns denoted by diagonal stripes, in 5,000 trading simulations in which, given the vote allocation, a trade is selected randomly, following the estimated probabilities in Table 7. As in all simulations in the paper, at each time interval the probability of a trade occurring is calculated so as to replicate, on average, the observed number of trades in the treatment. The model simulations match the ordinal ranks of the different outcomes' frequencies, although they consistently overestimate the frequency of the Condorcet winner. Such overestimation, however, is mostly mechanical: the result of the relatively high probability of random trades, and the likelihood that such random trades leave outcomes unchanged. Because zero-gain trades result in non-Pivotal vote allocations, they make Pivot-trades impossible, and thus bias the simulations

<sup>44</sup>In our matrices, the fewer the changes in the resolution of the different issues, the higher the aggregate payoff.

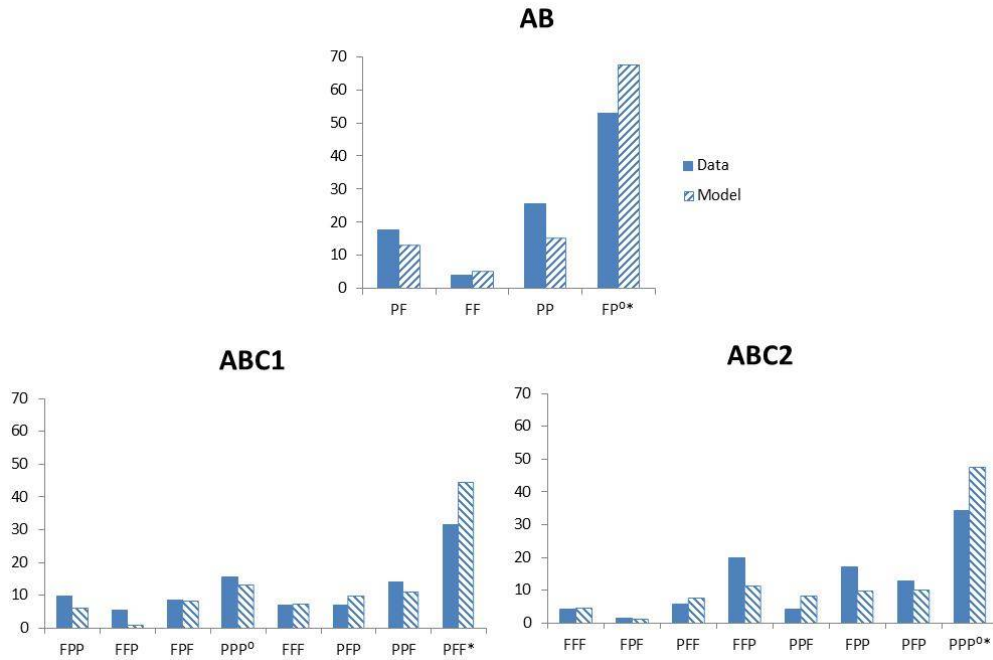


Figure 11: Simulated Outcomes

towards pre-trade outcomes and the Condorcet winner.

## 5 Conclusions

The aim of this paper is to develop a theoretical framework for studying vote trading in committees and to use the framework to help us understand data from a vote trading laboratory experiment. The theoretical framework has two essential features: (1) a notion of stability; and (2) a class of rational vote trading algorithms. With respect to (1), we define as stable those vote allocations for which there are no strict payoff-improving vote trades between any two voters. We then extend this definition to allow for coalitional trades. With respect to (2), we describe a family of algorithms called Pivot Algorithms, which define sequences of rational trades between pairs, or coalitions, of voters. We show that if trades are limited to pairs of voters, Pivot Algorithms always converge to a stable vote allocations. The result however breaks down for coalitional trades. We show that the outcomes reached through the algorithms have desirable welfare properties in special cases (when only two proposals are on the table, or when the committee consists of three voters and the Condorcet winner exists) but need not be desirable in general.

The experiment delivered three main findings. First, our notion of stability helps

in explaining the experimental data. Overall, two-thirds of the final vote allocations in the experiment are stable, and more than eighty percent are at most one trade away from stability. Second, although final vote allocations are in line with the theory, final outcomes show a clear bias towards the pre-trade outcome. In vote-trading environments, individual deviations from predicted trades can have large impacts. In particular, trades that cumulate votes on the side that is already winning make pivotal trades impossible and consolidate the pre-trade outcome. Finally, an analysis of trade-by-trade data lends weak support to the pivot algorithm itself. We classify trades through a simple statistical model and find that score-improving trades are more prevalent in the data than strict payoff-improving trades. Score-improving trades are vote exchanges in which each voter trades a vote on a less important issue in exchange for a vote on a more important issue, without necessarily benefiting from the trade (either because the outcome does not change, or because the directions of preferences are such that one voter suffers a loss). Because such trades co-exist with the subjects' ability to recognize stable vote allocations—i.e. to stop trading in the absence of further opportunities for payoff increases, we conjecture that they may be precautionary more than irrational. It is an intriguing conclusion, suggesting the possibility of farsighted behavior by the experimental subjects, in a very complex environment. We leave the discussion of such a possibility to future research.

## References

- [1] Bernholz, P. 1973, "Logrolling, Arrow Paradox and Cyclical Majorities", *Public Choice*, 15, 87-96.
- [2] Diamantoudi, E., E. Miyagawa, and L. Xue, 2004, "Random Paths to Stability in the Roommate Problem", *Games and Economic Behavior*, 48, 18-28.
- [3] Echenique F. and L. Yariv, 2012, "An Experimental Study of Decentralized Matching", mimeo, Caltech.
- [4] Ferejohn, J. 1974, "Some notes on the theory of vote trading", Caltech, Social Science Working Paper No.41.



- [5] Fischbacher U. and S. Schudy, 2014, "Reciprocity and Resistance to Comprehensive Reform", *Public Choice* 160:411-428.
- [6] Gale D. and L. Shapley, 1962, "College Admissions and the Stability of Marriage", *American Mathematical Monthly*, 69, 9-15.
- [7] Gusfield, D. and R. Irving, 1989, *The Stable Marriage Problem*, Cambridge Ma: MIT Press.
- [8] Jackson, M. and A. Watts, 2002, The Evolution of Social And Economic Networks, *Journal of Economic Theory*, 106(2): 265-295.
- [9] Jackson, M. and A. Wolinsky, 1996, A Strategic Model of Social and Economic Networks, *Journal of Economic Theory*, 71(1): 44E4.
- [10] Kadane, J., 1972, "On the Division of the Question", *Public Choice*, 13, 47-54.
- [11] Kagel, J. and A. Roth, 2000, "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment", *Quarterly Journal of Economics*, 115, 201-235.
- [12] Koehler, 1975, "Vote Trading and the Voting Paradox: A Proof of Logical Equivalence", *American Political Science Review*, 69, 954-960.
- [13] Ledyard, J.O., D. Porter, and A. Rangel, 1994, "Using Computerized Exchange Systems to Solve an Allocation Problem in Project Management", *Journal of Organizational Computing*, 4, 271-296.
- [14] McKelvey, R. and P. Ordeshook, 1980, "Vote trading: An experimental study", *Public Choice*, 35, 151-184.
- [15] Miller, N., 1977a, "Graph-Theoretical Approaches to the Theory of Voting", *American Journal of Political Science*, 769-803.
- [16] Miller, N., 1977b, "Logrolling, Vote Trading and the Paradox of Voting: A Game Theoretical Overview", *Public Choice*, 30, 51-75.
- [17] Mueller, D., 1967, "The Possibility of a Social Welfare Function. Comment", *American Economic Review*, 57, 1304-1311.

- [18] Nalbantian, H. and A. Schotter, 1995, "Matching and Efficiency in the Baseball Free-Agent System: An Experimental Examination", *Journal of Labor Economics*, 13, 1-31.
- [19] Niederle, M. and L.Yariv, 2011, "Decentralized Matching with Aligned Preferences", unpublished.
- [20] Pais,J., A. Pinter and R. Veszteg, 2011, "Decentralized Matching Markets: A Laboratory Experiment", mimeo, ISEG/Technical University of Lisbon.
- [21] Park, R.E., 1967, "The Possibility of a Social Welfare Function. Comment", *American Economic Review*, 57, 1300-1304.
- [22] Riker, W. and S. Brams, 1973, "The Paradox of Vote Trading", *American Political Science Review*, 67, 1235-1247.
- [23] Roth, A. and M. Sotomayor, 1990, *Two-Sided Matching: A Study in Game-Theoretic Modelling and Analysis*, Cambridge, UK: Cambridge University Press.
- [24] Roth A. and J. Vande Vate, 1990, "Random Paths to Stability in Two-Sided Matching", *Econometrica*, 58, 1475-1480.
- [25] Sasaki, H. and M. Toda, 1996, "Two-Sided Matching Problems with Externalities", *Journal of Economic Theory*, 70, 93-108.
- [26] Schwartz, T., 1975, "Vote Trading and Pareto Efficiency", *Public Choice*, 24, 101-110.
- [27] Stratmann, T. , 1992, "The Effects of Logrolling on Congressional Voting", *American Economic Review*, 82, 1162-1176.
- [28] Wilson, R., 1969, "An Axiomatic Model of Logrolling", *American Economic Review*, 59, 331-341.

## Appendix 1. For online Publication

**Proposition 2.** *If  $K = 2$ , then, for all  $N$ ,  $Z$ , and  $R$ : (1)  $\mathbf{P}(V_T)$  is unique. (2)  $\mathbf{P}(V_T)$  is Pareto optimal. (3) If a Condorcet winner exists, then  $\mathbf{P}(V_T)$  is the Condorcet winner. (4)  $\mathbf{P}(V_T)$  can never be the Condorcet loser. (5) If  $\mathbf{P}(V_T) \neq \mathbf{P}(V_0)$ , then a majority prefers  $\mathbf{P}(V_T)$  to  $\mathbf{P}(V_0)$ .*

**Proof.** (1). At any step  $t$ , if there is a unique blocking pair, then  $V_{t+1}$  is determined uniquely and to any unique vote allocation corresponds a unique outcome. Multiplicity can arise only if at any step there are multiple blocking pairs (possibly with some voters belonging to more than one pair). But if  $K = 2$  any trade by any of the blocking pairs must induce the identical change in outcome. Thus at any step the outcome of the vote is unique. And since a Pivot-stable outcome exists, it must be unique. Note that the Pivot-stable vote allocation need not be unique.

(2) With  $K = 2$ , there are four possible outcomes:  $\{\{\emptyset\}, \{1\}, \{2\}, \{1, 2\}\}$ . Suppose, with no loss of generality, that  $\mathbf{P}(V_0) = \{1, 2\}$ . With  $K = 2$ ,  $\mathbf{P}(V_T) \in \{\{1, 2\}, \{\emptyset\}\}$ . Majority voting at  $V_0$  implies that at least one trader must rank  $\mathbf{P}(V_0)$  strictly above all other outcomes. But if trade occurs and  $V_0$  is blocked, there must be at least two voters who strictly prefer  $\{\emptyset\}$  to  $\mathbf{P}(V_0)$ . One of those voters must have preferences:  $\{2\} \succ \{\emptyset\} \succ \{1, 2\} \succ \{1\}$ , and the other  $\{1\} \succ \{\emptyset\} \succ \{1, 2\} \succ \{2\}$ . Thus, if  $V_0$  is blocked, no outcome is Pareto-dominated and all are Pareto-optimal. If  $V_0$  is not blocked,  $\mathbf{P}(V_0) = \mathbf{P}(V_T)$  by definition, and again it is Pareto-optimal.

(3) Again suppose, with no loss, that  $\mathbf{P}(V_0) = \{1, 2\}$ . There are four possible preference types, which we call  $T_{1,2}, T_\emptyset, T_1$ , and  $T_2$ , with rankings as in Table 8. Preferences are transitive and the outcome in any cell is preferred to the outcome(s) below it; two outcomes are in the same cell when either could be ranked above the other.

Call  $n_{1,2}$  the number of voters of type  $T_{1,2}$ , and similarly for the other types. With  $\mathbf{P}(V_0) = \{1, 2\}$ , both proposals are supported by a majority of voters. Call  $M_1$  ( $M_2$ ) the number of voters who prefer 1 (2) to pass rather than fail. Then  $M_1 = n_{1,2} + n_1 \geq (N + 1)/2$ , and  $M_2 = n_{1,2} + n_2 \geq (N + 1)/2$ . We begin by asking whether any of the four possible outcomes could be the Condorcet winner. With  $\mathbf{P}(V_0) = \{1, 2\}$ , it follows immediately that  $\{1, 2\}$  is preferred by the majority to

$T_{1,2}$	$T_\emptyset$	$T_1$	$T_2$
1, 2	$\emptyset$	1	2
2	2	$\emptyset$	$\emptyset$
1	1	1, 2	1, 2
$\emptyset$	1, 2	2	1

Table 8:  $K = 2$ , possible preference types

$\{P^1\}$  and to  $\{2\}$ —more generally the no-trade outcome must be majority preferred to any outcome that differs in the direction in which a single proposal is decided. This is Park (1967) and Kadane’s (1972) result: if there is a Condorcet winner, it can only be the no-trade outcome. Here that is  $\mathbf{P}(V_0) = \{1, 2\}$ . Suppose then that  $\mathbf{P}(V_0)$  is the Condorcet winner.

If either  $M_1 > (N + 1)/2$  or  $M_2 > (N + 1)/2$ , no voter is pivotal,  $V_0$  cannot be blocked, and  $\mathbf{P}(V_T)$  trivially equals  $\mathbf{P}(V_0)$ , the Condorcet winner. Hence the proposition holds. The more interesting case is when  $M_1 = M_2 = (N + 1)/2$ ,  $V_0$  is blocked, and trade takes place. Notice that in such a case  $(N + 1)/2 = n_{1,2} + n_1 = n_{1,2} + n_2$ . Hence  $n_1 = n_2$ . Call such a number  $m$ .  $\{1, 2\}$  ( $\mathbf{P}(V_0)$ ) is the Condorcet winner if it beats  $\{\emptyset\}$  but the ranking of the two outcomes by types  $T_1$  and  $T_2$  is ambiguous. Call  $m_1^{PP}$  ( $m_1^\emptyset$ ) the number of voters of type  $T_1$  who rank  $\{1, 2\}$  above (below)  $\{\emptyset\}$ , and similarly for voters of type  $T_2$ .  $\mathbf{P}(V_0) = \{1, 2\}$  beats  $\{\emptyset\}$  if and only if  $(n + 1)/2 + m_1^{PP} + m_2^{PP} > (n - 1)/2 + m_1^\emptyset + m_2^\emptyset$ , or:

$$1 > (m_1^\emptyset + m_2^\emptyset) - (m_1^{PP} + m_2^{PP}) \quad (1)$$

Note that  $(m_1^\emptyset + m_2^\emptyset) + (m_1^{PP} + m_2^{PP}) \equiv 2m$ , an even number. But if the sum of two numbers is even, the difference of those two numbers is also even. Hence  $\{1, 2\}$  is the Condorcet winner if and only if  $(m_1^\emptyset + m_2^\emptyset) - (m_1^{PP} + m_2^{PP}) \equiv 2R$ , where  $R$  is an integer strictly larger than 0.

We now show that if  $\{1, 2\}$  is the Condorcet winner, then it must also be the Pivot-stable outcome. Any pair of traders blocking  $V_0$  must be such that one of them is counted in  $m_1^\emptyset$  and one is counted in  $m_2^\emptyset$ . With  $K = 2$ , voters can only trade votes once. Hence  $V_0$  is blocked once if  $\min(m_1^\emptyset, m_2^\emptyset) = 1$  and any  $V_t$  such that  $\mathbf{P}_t = \{1, 2\}$  can potentially be blocked  $s$  times if  $\min(m_1^\emptyset, m_2^\emptyset) = s$ . Similarly, if  $V_0$  is blocked and

$V_1$  is such that  $\mathbf{P}_1 = \{\emptyset\}$ ,  $V_1$  will be blocked if  $\min(m_1^{PP}, m_2^{PP}) = 1$ . As above, any  $V_t$  such that  $\mathbf{P}_t = \{\emptyset\}$  can potentially be blocked  $s'$  times if  $\min(m_1^{PP}, m_2^{PP}) = s'$ . Thus  $\{1, 2\}$  is the Pivot-stable outcome if and only if  $\min(m_1^{PP}, m_2^{PP}) \geq \min(m_1^\emptyset, m_2^\emptyset)$ . Now recall that  $(m_1^\emptyset + m_1^{PP}) = (m_2^\emptyset + m_2^{PP})$  since both sums must equal  $m$ . Thus if  $\{1, 2\}$  is the Condorcet winner and  $(m_1^\emptyset + m_2^\emptyset) = (m_1^{PP} + m_2^{PP}) - 2R$ , it must be that  $m_1^\emptyset = m_2^{PP} - R$  and  $m_2^\emptyset = m_1^{PP} - R$ . Hence  $\min(m_1^{PP}, m_2^{PP}) = \min(m_1^\emptyset, m_2^\emptyset) + R$ , or  $\min(m_1^{PP}, m_2^{PP}) > \min(m_1^\emptyset, m_2^\emptyset)$ , and  $\{1, 2\}$  is the Pivot-stable outcome. Note that identifying  $\mathbf{P}(V_0)$  with  $\{1, 2\}$  is with no loss of generality. The proof can be restated as follows: the only candidate for Condorcet winner is  $\mathbf{P}(V_0)$ , and when  $\mathbf{P}(V_0)$  is the Condorcet winner, then it must also be the Pivot-stable outcome.

(4) By Kadane's "improvement algorithm", if  $\mathbf{P}(V_0) = \{1, 2\}$ , then not only is  $\{1, 2\}$  majority preferred to  $\{1\}$  and to  $\{2\}$ , but  $\{1\}$  and  $\{2\}$  are majority preferred to  $\{\emptyset\}$ . Hence if there is a Condorcet loser, it can only be  $\{\emptyset\}$ . But if  $\{\emptyset\}$  is the Condorcet loser, it means that  $\{1, 2\}$  is majority preferred to  $\{\emptyset\}$ . Hence  $\mathbf{P}(V_0) = \{1, 2\}$  is the Condorcet winner, and by (2) above  $\mathbf{P}(V_T) = \mathbf{P}(V_0)$ .

(5). From (2) above, if  $\mathbf{P}(V_0)$  is the Condorcet winner, then  $\mathbf{P}(V_T) = \mathbf{P}(V_0)$ . Thus if  $\mathbf{P}(V_T) \neq \mathbf{P}(V_0)$ ,  $\mathbf{P}(V_0)$  is not the Condorcet winner. If, with no loss of generality,  $\mathbf{P}(V_0) = \{1, 2\}$ , then if  $\mathbf{P}(V_T) \neq \mathbf{P}(V_0)$ ,  $\mathbf{P}(V_T) = \{\emptyset\}$ . By Kadane's argument,  $\mathbf{P}(V_0) = \{1, 2\}$  is majority preferred to  $\{1\}$  and to  $\{2\}$ . Hence if  $\mathbf{P}(V_0)$  is not the Condorcet winner,  $\{\emptyset\}$  must be majority preferred to  $\{1, 2\} = \mathbf{P}(V_0)$ . But  $\{\emptyset\} = \mathbf{P}(V_T)$ . Hence if  $\mathbf{P}(V_T) \neq \mathbf{P}(V_0)$ ,  $\mathbf{P}(V_T)$  must be majority preferred to  $\mathbf{P}(V_0)$ .  $\square$

**Proposition 3.** *If  $N = 3$ , then for all  $K$ ,  $Z$ , and  $R$ : (1) if a Condorcet winner exists,  $\mathbf{P}(V_T)$  is unique and is the Condorcet winner. (2)  $\mathbf{P}(V_T)$  can never be the Condorcet loser.<sup>45</sup>*

**Proof.** (1). Select any  $k$  proposals, with  $k = 2, \dots, K - 1$ . Call  $\mathbf{P}(V_0, k^-)$  the outcome that would follow if the  $k$  proposals were decided against the majority preference at  $V_0$ , and the remaining  $K - k$  according to the majority preference at  $V_0$ , i.e. as if one vote trade was executed on the  $k$  proposals and none on the remainder. If the Condorcet winner exists, by Park and Kadane, it can only be  $\mathbf{P}(V_0)$ . Thus for any  $\mathbf{P}(V_0, k^-)$  at least two of the three voters prefer  $\mathbf{P}(V_0)$  to

<sup>45</sup>Result (1) replicates the result in Koehler (1975), under slightly different assumptions. We find however that the result is limited to  $N = 3$ .

$\mathbf{P}(V_0, k^-)$ . But then no trade can take place. If the Condorcet winner exists,  $V_0$  cannot be blocked. Thus  $\mathbf{P}(V_T)$  equals  $\mathbf{P}(V_0)$  and is the Condorcet winner.

(2) We begin by reiterating, and generalizing, an argument we used above in the proof of Proposition 2.

**Lemma 2.** *Call  $\mathbf{P}_-(V_0)$  the outcome obtained by choosing the minority's preferred direction for each proposal at  $V_0$ . If a Condorcet loser exists, it can only be  $\mathbf{P}_-(V_0)$ .*

**Proof of Lemma 2.** The Lemma follows from Kadane's improvement algorithm. Select  $k$  from the  $K$  proposals ( $k = 1, \dots, K - 1$ ). Consider the outcome  $\mathbf{P}(V_0, k^-)$  obtained by deciding those  $k$  proposals in the direction favored by the minority at  $V_0$ , and the remainder  $K - k$  in the direction favored by the majority. Now consider the outcome obtained by switching one additional proposal from the majority to the minority's preferred direction at  $V_0$ :  $\mathbf{P}(V_0, k^-, (k + 1)^-)$ , where the argument  $k^-$  is maintained to make clear that the selection of the original  $k$  proposals has not changed. Then, by construction,  $\mathbf{P}(V_0, k^-)$  is majority preferred to  $\mathbf{P}(V_0, k^-, (k + 1)^-)$ . It follows that for any  $\mathbf{P} \neq \mathbf{P}_-(V_0)$  there always exists an outcome  $\mathbf{P}'$  that differs only by switching the direction of one proposal from the majority's to the minority's at  $V_0$  such that  $\mathbf{P}$  is majority preferred to  $\mathbf{P}'$ . Hence if a Condorcet loser exists, it can only be  $\mathbf{P}_-(V_0)$ .  $\square$

We can now prove (2). By Lemma 2, if a Condorcet loser exists it can only be  $\mathbf{P}_-(V_0)$ . The result follows if we can show that if  $\mathbf{P}_-(V_0)$  is Pivot-stable, it cannot be the Condorcet loser. Suppose  $\mathbf{P}_-(V_0)$  is Pivot-stable, and call  $V_0$  any allocation of votes such that  $\mathbf{P}(V_0) = \mathbf{P}_-(V_0)$ . Given preferences  $Z$ , consider any allocation of votes  $V(Z)$  such that  $V_0$  can be reached from  $V(Z)$  through a strictly pair-wise improving trade. Since  $\mathbf{P}_-(V_0) \neq \mathbf{P}(V_0)$  and  $\mathbf{P}_-(V_0)$  is Pivot-stable,  $V(Z)$  must exist. With  $N = 3$ , the existence of a strictly pair-wise improving trade implies that  $\mathbf{P}(V_0) = \mathbf{P}_-(V_0)$  is majority preferred to  $\mathbf{P}(V(Z))$ . Hence if  $\mathbf{P}_-(V_0)$  is Pivot-stable, it cannot be the Condorcet loser, and the result is proven.  $\square$

**Proposition 5.** *There exist  $K$ ,  $N$ ,  $Z$ , and  $R_C$  such the C-Pivot algorithm never reaches a stable vote allocation.*

**Proof.** Consider the following example, where as usual rows represent proposals, columns represent voters and the entry in each cell is  $z_i^k$ , the value attached by voter

	1	2	3	4	5	6	7
A	3	-2	-2	-2	1	1	1
B	-2	3	-2	-2	1	1	1
C	-2	-2	3	-2	1	1	1
D	-2	-2	-2	3	1	1	1

Table 9: *Profile where C-Pivot never converges*

$i$  to proposal  $k$  passing.

At  $V_0$ , all proposals pass, and  $u_i(V_0) = -3$  for  $i = \{1, 2, 3, 4\}$ . Consider a coalition composed of such voters, and the following coalition trade: voter 1 gives his  $A$  vote to voter 2, in exchange for his  $B$  vote; voter 3 gives his  $C$  vote to voter 4, in exchange for his  $D$  vote. At  $V_1$ , all proposals fail and  $u_i(V_1) = 0$  for all  $i \in C$ . The trade is strictly improving for all members of the coalition. In addition, it is a minimal trade, since  $V_1$  cannot be reached by the coalition by trading fewer votes, nor can it be reached by a smaller coalition (each of the two pair-wise trades alone is welfare decreasing for the pair involved). But note that  $V_1$  is not C-Pivot stable: voters 1 and 2 can block  $V_1$  by trading back their respective votes on  $A$  and  $B$ , reaching outcome  $\mathbf{P}(V_2) = \{A, B\}$ , and enjoying a strictly positive increase in payoffs:  $u_j(V_2) = 1$  for  $j = \{1, 2\}$ . The same argument applies to voters 3 and 4. At  $V_2$ ,  $u_s(V_2) = -4$  for  $s = \{3, 4\}$ , but 3 and 4 can block  $V_2$ , trade back their votes on  $C$  and  $D$ , and obtain a strict improvement in their payoff:  $\mathbf{P}(V_3) = \{A, B, C, D\}$ , and  $u_s(V_3) = -3$  for  $s = \{3, 4\}$ . Note that the sequence of trades has generated a cycle:  $V_3 = V_0$ , an allocation that is blocked by coalition  $C = \{1, 2, 3, 4\}$ , etc.. Hence for  $R_C$  that selects the blocking coalitions in the order described, no C-Pivot stable allocation of votes can be reached.  $\square$

The logic we exploited in the proof of the Theorem in section 3 does not extend to coalition trades. Because of the externalities present in coalitional trades, the score function we defined earlier is no longer monotonically increasing in the number of trades. In the example, voters 1, 2, 3, and 4 have a score of 9 before the coalition trade and a score of 8 after the trade.<sup>46</sup> Cycles become possible, and stability cannot

<sup>46</sup>In the example, the coalition trade can be divided into two separate pair-wise trades, but this feature plays no important role. It is easy to generate examples where coalition trades are linked in a chain. What matters is that scores can fall after a welfare-improving coalition trade.

be guaranteed.

**Proposition 7.** *Consider  $Z$  such that the Condorcet winner exists. (1) If either  $K = 2$  or  $N = 3$ , then the C-Pivot stable outcome always exists, is unique, and coincides with the Condorcet winner. (2) If  $K > 2$  and  $N > 3$ , if a C-Pivot stable outcome exists, it need not coincide with the Condorcet winner.*

**Proof.** (1) (i) Consider first the case  $K = 2$ . The following Lemma establishes that the existence of a C-Pivot stable outcome:

**Lemma 3.** *If  $K = 2$ , then for all  $N$ ,  $Z$ , and  $R$  a C-Pivot stable allocation  $V$  always exists.*

**Proof of Lemma 3.** We need two further Lemmas:

**Lemma 4.** *If at  $V_0$  both proposals pass by minimal majority, then if  $K = 2$  all C-Pivot trades must be pair-wise trades.*

**Proof of Lemma 4.** By minimality of the coalition trades, if at  $V_0$  both proposals pass by minimal majority then at each step of the algorithm both proposals must be decided by minimal majority. At each step  $t$  any coalition must include a voter with preferences  $\{1\} \succ \mathbf{P}(V_t)$  and a voter with preferences  $\{2\} \succ \mathbf{P}(V_t)$ , and each must have at least one vote to trade away, or they would not be part of the minimal coalition. But if the proposals are decided by minimal majority, the two voters could just trade among themselves. Hence the minimal coalition must be a pair.  $\square$

**Lemma 5.** *If  $K = 2$ , there can be at most one trade that is not pair-wise, and it can only be the first.*

**Proof of Lemma 5.** By Lemma 4, if non pair-wise coalition trades occur, it must be that at  $V_0$  at least one proposal is not decided by minimal majority. Any blocking coalition at  $V_0$  must then involve more than two voters. But by minimality, after the coalition trade both proposals must be decided by minimal majority. But then from  $t = 1$  onward, all trades can only be pair-wise trades.  $\square$

We can now prove Lemma 3. If  $V_0$  cannot be blocked, then it is trivially stable. If it can be blocked, then Lemma 4 and 5 imply that any voter can at most trade once. Hence the trades must converge to a stable votes allocation. We can say more: a stable votes allocation must be reached in at most  $(N - 1)/2$  steps, the maximal



number of possible pair-wise trades.  $\square$

We can now prove that the  $C$ -Pivot stable outcome must coincide with the Condorcet winner, when the Condorcet winner exists, and thus be unique.<sup>47</sup> If at  $V_0$  both proposals pass by minimal majority then by Lemma 4 Proposition 2 applies and the result follows. Suppose then that at  $V_0$  at least one proposal is not decided by minimal majority. If there is no blocking coalition,  $V_0$  is trivially stable, and since the Condorcet winner can only be  $\mathbf{P}(V_0)$ , then the stable outcome coincides with the Condorcet winner, if it exists. Suppose then that at least one blocking coalition exists (and note that it must include more than two voters). If several exist, select one by rule  $R_C$ . After the first coalitional trade,  $\mathbf{P}(V_1) = \{\emptyset\}$ . With  $K = 2$ , all members of a minimal coalition can only trade once, regardless of the coalition's size. We can construct a fictional vote allocation  $\tilde{V}_0$  and preferences  $\tilde{Z}$  such that  $\tilde{v}_{0i}^k = v_{1i}^k$  and  $\tilde{z}_i^k = z_i^k$  for all  $i \notin C$ , and  $\tilde{v}_{0i}^k = 1$  and  $\tilde{z}_i^k = -1$  for all  $i \in C$ ,  $k = 1, 2$ . Note that  $\mathbf{P}(\tilde{V}_0) = \mathbf{P}(V_1) = \emptyset$ , and  $\tilde{Z}$  respects all individuals' ranking between the only possible outcomes,  $\{1, 2\}$  and  $\{\emptyset\}$ , but  $\tilde{V}_0$  and  $\tilde{Z}$  guarantee that, as required, all  $i \in C$  will not trade any further. At  $\tilde{V}_0$ , both proposals are decided by minimal majority, and starting from  $\tilde{V}_0$  all minimal coalitions will be pair-wise. As a result, starting from  $\tilde{V}_0$ , Proposition 2 applies. Note that, if  $\mathbf{P}(V_0) = \{1, 2\}$  is the Condorcet winner, then a majority prefers  $\mathbf{P}(V_0)$  to  $\mathbf{P}(\tilde{V}_0)$ . Hence, by Proposition 2,  $\mathbf{P}(\tilde{V}_0)$  cannot be Pivot stable. The  $C$ -Pivot stable outcome is then  $\mathbf{P}(V_0)$ . But  $\mathbf{P}(V_0)$  is the Condorcet winner and the result is proven.

(ii). Consider now the case  $N = 3$ . Recall that the Condorcet winner can only be  $\mathbf{P}(V_0)$ . If  $N = 3$ , and the Condorcet winner exists, then no welfare improving trade exists at  $V_0$ . Hence  $\mathbf{P}(V_0)$  is trivially  $C$ -Pivot stable, and is the Condorcet winner.  $\square$

### Trade paths for the experimental matrices.

(1) Matrix  $AB$ . There is a unique Pivot path. At  $V_0$ , with  $\mathbf{P}(V_0) = \{B\}$ , 1 trades his  $B$  vote to 3 in exchange for 3's  $A$  vote, leading to  $\mathbf{P}(V_1) = \{A\}$ ; then 2 trades his  $B$  vote to 4 in exchange for 4's  $A$  vote. The resulting vote allocation is stable, and  $\mathbf{P}(V_2) = \mathbf{P}(V_T) = \{B\}$ . (2) Matrix  $ABCL$ .  $\mathbf{P}(V_0) = \{A\}$  is

<sup>47</sup>In fact, with  $K = 2$  it is possible to show that the  $C$ -Pivot stable outcome is always unique, whether or not the Condorcet winner exists, and it is always Pareto optimal and always majority preferred to  $\mathbf{P}(V_0)$ .

the Condorcet winner, but  $V_0$  is not stable. The unique Pivot-stable outcome is  $\mathbf{P}(V_T) = \{A, B, C\}$ . Three alternative paths, of length  $\{4, 4, 3\}$  lead to it. Indicating first the ID's of the trading partners, and then, in lower-case letters, the issue on which an extra vote is acquired by the voter listed first, the three paths are:  $\{13cb, 45bc, 23ab, 45ca\}$ ,  $\{23ab, 45ca, 45bc, 13cb\}$ , and  $\{23ab, 45ba, 13cb\}$ . (3) Matrix  $ABC^2$ .  $\mathbf{P}(V_0) = \{A, B, C\}$  is the Condorcet winner and the unique Pivot-stable outcome. However  $V_0$  is not stable. Three alternative paths, of length  $\{4, 4, 3\}$  lead to  $\mathbf{P}(V_T) = \{A, B, C\}$ . They are:  $\{15ab, 34ba, 24cb, 15bc\}$ ,  $\{24cb, 15bc, 15ab, 34ba\}$ , and  $\{24cb, 15ac, 34ba\}$ .

## Appendix 2. For online publication

### VOTE TRADING INSTRUCTIONS

Make yourself comfortable, and then please turn off phones and don't talk or use the computer. Thank you for agreeing to participate in this decision making experiment. You will be paid for your participation in cash, at the end of the experiment. Different participants may earn different amounts. What you earn depends partly on your decisions and partly on the decisions of others. If you have any questions during the instructions, raise your hand and your question will be answered. If you have any questions after the experiment has begun, raise your hand and an experimenter will come and assist you.

The experiment today is a committee voting experiment, where you will have an opportunity to trade votes before voting on an outcome. The experiment will be in three parts. At the end of the experiment you will be paid the sum of what you have earned in all three parts of the experiment, plus your promised show-up fee of 10 dollars. Everyone will be paid in private and you are under no obligation to tell others how much you earned. Your earnings during the experiment are denominated in POINTS. For this experiment every 100 POINTS earns you 6 DOLLARS.

Here are the instructions for Part I.

You will be randomly assigned to one of 3 committees, each composed of 5 members. Each committee is completely independent of the others, and the decision taken in one committee has no effect on the others. The committee will vote using majority rule to decide on 2 different motions, denoted A and B. Each motion can either pass or fail depending on how the committee votes. There will be a separate vote on each motion. The computer will assign you a committee member number (1, 2, 3, 4, or 5). Part I consists of 3 rounds.

You will be told, for each motion, whether you prefer it to pass or to fail. The computer will assign you (and each other member) a value for each motion which will be a number between 1 and 100. You will earn your value for a motion if you prefer that motion to pass and it passes, or if you prefer it to fail and it fails. This is your only source of earnings. Your earnings for the round are equal to the sum of your earnings over the two motions.

Each committee member starts a round with 1 vote to cast on each motion. Then

there will be a 2 minute trading period, during which you and the other members of your committee will have an opportunity to trade votes with each other. For example, you may wish to trade your A vote in exchange for some other member's B vote. We will describe exactly how to do this shortly.

After the trading period ends, you will proceed to the voting stage. Once every one has voted, you will be told what the final votes were in your committee and how much you earned in that round. This will complete the first round. The remaining 2 rounds in Part 1 follow the same rules. Each committee member starts the round with a single vote on each motion. Your committee member number, preferences for each motion (pass or fail), your value for each motion, and the preferences and values of the other four members of your committee all stay the same for all 3 rounds of part 1 of the experiment.

Your earnings for part 1 are the sum of your earnings in all 3 rounds. After round 3 ends, I will read you instructions for part 2 of the experiment.

We now describe in detail how you and the other members of your committee can trade votes. When we begin a round, you will see a screen like this, although the exact numbers may be different. [Display Screen 1] On the right of the screen is an Information Table that contains a lot of information, so please listen carefully. It displays each member's preference for each motion (pass or fail), value, and number of votes. If the member prefers the motion to fail, then the value is written in a blue color. If the member prefers the motion to pass, then the value is written in an orange color. You can simply think of there being two sides - the orange side and the blue side - on each motion. The number of votes held by each member on each motion is in parentheses. Because no trading has occurred yet, each member holds exactly one vote on each motion.

Your own row is specifically labeled and the label is highlighted in gray. The last row in the table is labeled "outcome". This row tells you, for each motion, what the total vote would be if voting took place now, by showing the column sum of votes on each motion. The number of votes for is given first, in orange, and the number of votes against is given second, in blue. If the votes in favor of a motion exceed the votes against, then all voters who prefer the motion to pass will earn their value for that motion, and all voters who prefer the motion to fail will earn zero for that motion. Similarly, if the votes in favor of a motion failing exceed the votes in favor

of it passing, then all voters who prefer the motion to fail will earn their value for that motion, and all voters who prefer the motion to pass will earn zero for that motion. There is a check mark next to your value if the outcome of that motion is the outcome you prefer. This means that you earn your value for that motion. In this example, if there were no votes traded at all, then on motion A, there are 2 votes held by members who prefer A to pass and 3 held by members who prefer A to fail, so motion A fails. On motion B, there are 3 votes held by members who prefer B to pass and 2 held by members who prefer B to fail, so motion B passes. Since ID 1 (You) prefers both motions to pass, he earns his value for motion B but earns 0 for motion A.

To the left of the table, in grey, is the trading window. At any time during the trading period, any committee member may post a trade offer by requesting 1 vote on one motion in exchange for 1 vote on some other motion. Suppose the participant on the slide in front of the room wanted to post a trade requesting one A vote in exchange for one B vote. This is done by entering a 1 in the A box under "Requests" and a 1 in the B box under "Offers". [Screen 2]. You can only trade 1 vote for 1 vote; you can neither request nor offer multiple votes.

After you have entered this trade request and clicked the "submit trade offer" button, the trade is posted in the trading panel for everyone in your committee to see. [SCREEN 3] If another committee member wants to accept your trade request, they may click on it to highlight it, and then click on the "accept selected offer" button.[SCREEN 4] You now have 10 seconds to either confirm or reject the accepted trade. A message will pop-up on your screen. [SCREEN 5]. The message tells you what the outcome of the vote would be if you either accept or reject the trade and voting took place without any further trade. If you reject the trade or do nothing for 10 seconds, the trade does not occur. The committee member who had accepted your offer is informed that you declined to confirm the trade.[SCREEN 6]. Your offer is reposted in the trading window, and some other voter can accept it. If you confirm the trade, then the voter who accepted the offer now holds 0 A votes and 2 B votes, you now hold 2 A votes and 0 B votes, and the Information Table is updated accordingly. The new Information Table is displayed for 10 seconds on a popup screen for everyone in your group to see. [SCREEN 7]

If you have a standing offer listed in the trading window, you may cancel it by

first clicking on it and then clicking the "cancel selected offer" button. [SCREEN 8]

The trading period continues for 2 minutes. The timer at the top tells you how much time remains in the trading period. The clock is frozen when the Information Table is shown after a trade, with the new vote holdings. If a trade occurs within 10 seconds of the end of the trading period, the trading period is automatically lengthened by 10 more seconds.

You are free to post trade requests at any time, but you are not allowed to offer to trade away a vote on a motion if you currently hold 0 votes for that motion or already have an offer posted on the trading window that would result in holding 0 votes if accepted. In that case you would first have to cancel your existing posted offer. Also remember that you can only trade one vote for one motion in exchange for one vote for another motion. If you try to do a trade that is not allowed, you will either receive an error message, or the action buttons will become gray and be deactivated, preventing you from proceeding with that trade.

When the trading period for the round is over, we proceed to the voting stage. Your screen will now look something like this: [SCREEN 9]. In this stage you do not really have any choice. You are simply asked to click a button to cast all the votes you hold at the end of trading. The computer will automatically cast your votes on each motion according to the preferences you were assigned. For example, if you prefer motion B to fail and you hold two B votes after the trading period, those two votes will be cast automatically against motion B. Please cast all your votes without delay by clicking on the vote button.

After you and the other members of the committee have voted, the results are displayed and summarized. [SCREEN 10]

As the experiment proceeds, at the bottom of each screen you will see a history table, summarizing the results of the previous rounds [SCREEN 11. Go over the different columns] If you switch to tab view, each round will be shown separately].

We then proceed to the next round, where you again start out with one vote on each motion and the rules are the same as in the first round. Remember that your assigned committee number, preferences for motions, values for motions, and those of the other members of your committee all stay the same for all 3 rounds of part 1 of the experiment. After the first 3 rounds are completed, we will read instructions for the second part of the experiment.

To give you some experience with the trading screen, we will conduct two practice rounds. The rules will be the same as they will be in the paid rounds, but the values and preference assignments, for or against a motion, are not the same as they will be in the paid rounds. You are not paid for the practice rounds, so they have no effect on your final earnings. The only purpose of the practice rounds is to help you become familiar with the computer interface and the trading rules.

This summary slide [SCREEN 12: Summary slide] will remain up during the experiment to remind you of the rules on trading and on time.

Are there any questions before we proceed to the first practice round? [START SERVER]

Please click on the icon marked Multistage Client on your desktop. Then enter the number of your carrel (on the right side of the carrel), click enter, and then wait. Remember that you are not allowed to use the computer for any other purposes while waiting during the experiment (email, browsing, etc.).

[CONNECT EVERYONE AND START]

Please complete the practice rounds on your own. Feel free to raise your hand if you have a question.

[WAIT FOR SUBJECTS TO COMPLETE PRACTICE ROUNDS]

The practice rounds are now over. Remember, you will not be paid the earnings from the practice rounds.

If you have any questions from now on, raise your hand, and an experimenter will come and assist you. We will now begin the paid rounds.

(Play 3 real rounds for Part 1) [After last ROUND, read:]

We will now proceed to Part 2. The rules for part 2 are the same as for part 1, but there are now 3 motions for your group to vote on. You can only trade one vote on one motion for one vote on another motion. The trading period will last 3 minutes. As before, 10 seconds will be added to the clock if a trade takes place within 10 seconds of the time limit.

The values and pass/fail preferences will be different from part 1, and your committee number as well as the composition of your committee may change. However, both the preferences and the composition of the committee will remain the same for all of Part 2. Part 2 will last for 5 rounds. At the end of the 5 rounds, we will stop

and read the instructions for Part III.

Are there any questions before we begin?

(Play 5 real rounds for part 2) [After last ROUND, read:]

We will now proceed to Part 3. Part 3 is identical to Part 2, but the values and pass/fail preferences may be different. Your committee number as well as the composition of your committee may also change. Part 3 will again last for 5 rounds and again the trading period is 3 minutes (plus 10 seconds if a trade is concluded within 10 seconds of the time limit).

This is the end of the experiment. You should now see a popup window, which displays your total earnings in the experiment. Please record this and your Computer ID on your payment receipt sheet, rounding up to the nearest dollar. After you are done, please, click ok to close the popup window. Do not close any other windows on your computer and do not use your computer for anything else. Also enter 10 dollars on the show-up fee row. Add the two numbers and enter the sum as the total.

Write output

We will pay each of you in private in the next room in the order of your computer numbers. Remember you are under no obligation to reveal your earnings to the other players. Please do not use the computer; be patient, and remain seated until we call you to be paid. Do not converse with the other participants or use your cell phone. Thank you for your cooperation.



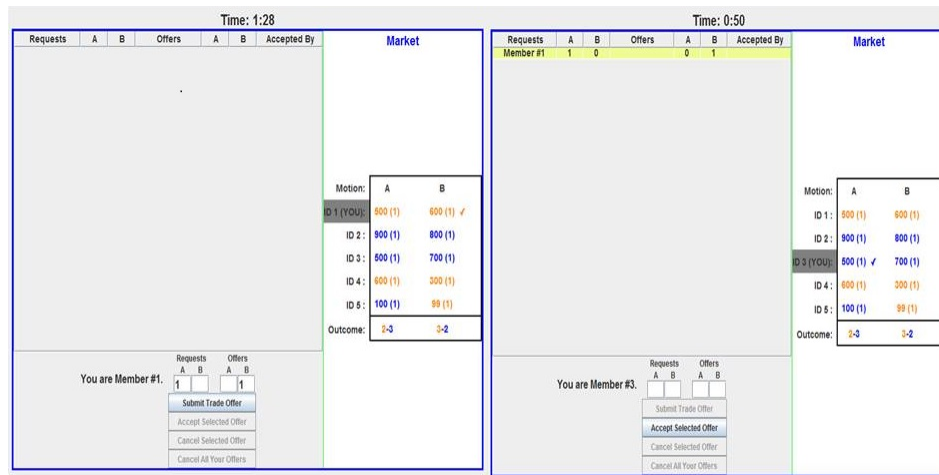


Figure 12: Screenshots for a subject posting a bid (on the left), and for a subject accepting a posted bid (on the right).

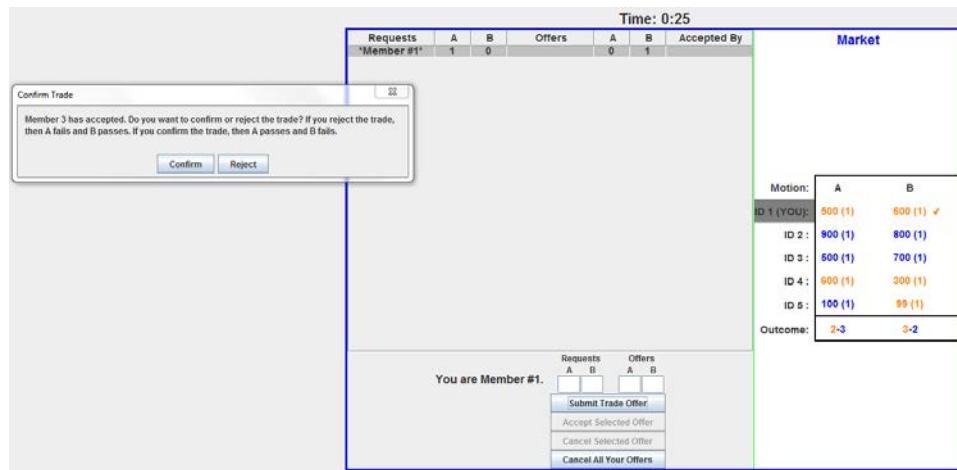


Figure 13: Confirmation request for the bidder.