Trading Votes for Votes. A Dynamic Theory[1]

Alessandra Casella[2] Thomas Palfrey[3]

October 10, 2018

[1]We thank Yimeng Li, Kirill Pogorelskiy, and Enrico Zanardo for research assistance, and participants at several research conferences and seminars for their comments. We especially thank Scott Ashworth, Micael Castanheira, Reyer Gerlagh, Michel LeBreton, Cesar Martinelli, Debraj Ray, Richard Van Weelden, Rajiv Vohra, and Alistair Wilson for detailed comments and suggestions. The paper has also benefitted from detailed comments from the editor and referees. The National Science Foundation (SES-1426560; SES-0617934) and the Gordon and Betty Moore Foundation (SES-1158) 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 present manuscript expands on an earlier working paper version entitled Trading Votes for Votes: A Decentralized Matching Algorithm.

[2]Columbia University, NBER and CEPR, ac186@columbia.edu

[3]Caltech and NBER, trp@hss.caltech.edu

Abstract

We develop a framework to study the dynamics of vote trading over multiple binary issues. We prove that there always exists a stable allocation of votes that is reachable in a finite number of trades, for any number of voters and issues, any separable preference profile, and any restrictions on the coalitions that may form. If at every step all blocking trades are chosen with positive probability, convergence to a stable allocation occurs in finite time with probability one. If coalitions are unrestricted, the outcome of vote trading must be Pareto optimal, but unless there are three voters or two issues, it need not correspond to the Condorcet winner.

JEL Classification: C62, C72, D70, D72, P16

Keywords: Voting, Vote Trading, Condorcet Winner, Stability

1 Introduction

Exchanging one's support of a proposal for someone else's support of a different proposal is common practice in group decision-making. Whether in small informal committees or in legislatures, common sense, anecdotes, and systematic evidence all suggest that vote trading is a routine component of collective decisions.[1] Vote trading is ubiquitous, and yet its theoretical properties are not well understood. Efforts at a theory were numerous and enthusiastic in the 1960's and 70's but fizzled and have almost entirely 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: each vote trade occurs without the equilibrating properties of a continuous price mechanism, causes externalities to allies and opponents of the trading parties, and can trigger new profitable exchanges. As a subset of voters trade votes on a set of proposals, the default outcomes of these proposals change in response to the reallocation of votes, generating incentives for a new round of vote trades, which will again change outcomes and open new trading possibilities. A second reason for the early difficulties is that a consistent well-defined framework was missing. Most authors left unspecified some crucial details of their models, used an array of different assumptions and terminology, at times implicit, and never fully closed the loop between the definition of stability and the specification of the trading rule. The first contribution of this paper is the development of a general theoretical framework for analyzing vote trading as a sequential dynamic process.

The voting environment is comprised of an odd number of voters facing several binary proposals, each of which will either pass or fail. Every committee member can be in favor or opposed to any proposal. Preferences are represented by intensities over winning any individual proposal and are additively separable across proposals, inducing for each voter an ordering over all possible outcomes – i.e., all combinations of different proposals passing or failing.

An initial allocation of votes specifies how many votes each voter controls on each proposal. After vote trades are concluded, each proposal is decided by majority rule: if, after trading, the number of votes controlled by voters favoring a proposal exceed the number of votes controlled by voters opposing a proposal, then the proposal passes; otherwise it fails.

Votes are tradable, in the sense of a barter market. A vote trade is a reallocation of votes held by a subset, or coalition, of voters. Hence the dynamic process operates on the set of feasible vote allocations, and the current state of the dynamic system corresponds to the current allocation of votes. We specify a family of simple algorithms, called Pivot algorithms, according to which trading evolves over time. Dynamic sequences of vote trades are executed by sequences of blocking

[1] An empirical literature in political science documents vote trading in legislatures. For example, Stratmann (1992) provides evidence of vote trading in agricultural bills in the US Congress.

coalitions–subsets of voters who, given the allocation of votes, can reallocate votes among themselves and reach a new outcome each of them strictly prefers to the pre-trade outcome. Both the coalition and the trade are fully unconstrained: the coalition can be of any size, and each member can trade as many votes as she wishes on as many proposals as desired; trades need not be one-to-one. The only requirement is that all members of a blocking coalition must strictly gain from the trade. If the initial vote allocation is not blocked by any coalition, it is stable, and there is no trade. However if the allocation is blocked, it may be blocked by many different coalitions and many different trades. An element of our family of algorithms is any rule selecting blocking coalitions and trades at each blocked allocation. The trade produces a new allocation of votes, and the algorithm again selects a blocking coalition and a trade. The algorithm continues until a vote allocation is reached from which there are no improving trades for any coalition. Such vote allocation is called Pivot stable.

The approach delivers three main results, addressing some of the open questions left from the older literature. Our first key result is a general existence theorem. The set of Pivot stable vote allocations is non-empty. For any initial vote allocation, any number of voters or proposals, any profile of preference rankings, any restrictions on feasible blocking coalitions, there always exists a finite sequence of trades that ends at a stable allocation. The existence result does not rule out the possibility that some selection rules may generate cycles. However, if every blocking trade is selected with positive probability, then trade must converge to a Pivot stable vote allocation in finite time with probability one. Furthermore, if trades are restricted to be pairwise and non-redundant–i.e., if votes that do not affect outcomes are not traded – then trading converges to a stable allocation along *all* possible sequences of blocking trades. Earlier conjectures (Riker and Brams, 1973; Ferejohn, 1974) speculated that vote trading could reach a stable allocation only under very strict conditions on the number and types of trades, and in particular ruling out coalitional trades. Our results show otherwise: a stable allocation is always reachable.

Every vote allocation produces an outcome, that is, a specific combination of proposals passing or failing. Our second result concerns the optimality of Pivot stable allocations, and serves as a welfare theorem to complement the existence theorem. Pivot stable vote allocations always generate Pareto optimal outcomes if no restrictions are placed on the set of blocking coalitions. Together with our existence result, we can then conclude that vote trading can always deliver a stable Pareto optimal outcome.

The early literature was inspired in large part by a claim, stated explicitly in Buchanan and Tullock (1962), that vote trading must lead to Pareto superior outcomes because it allows the expression of voters' intensity of preferences.[2] The conjecture was rejected by Riker and Brams' (1973) influential "paradox of vote trading" which showed that when trade is restricted to be pairwise, Pareto-inferior outcomes are possible. The belief that constraining trade to be pairwise was necessary to achieve stability made the conclusion particularly important. Our result is consistent with the Riker and Brams' paradox because restrictions on the allowable set of blocking coalitions

[2] The claim originated in an early debate between Gordon Tullock and Anthony Downs (Tullock, 1959 and 1961, Downs, 1957, 1961). See also Coleman (1966), Haefele (1971), Tullock (1970), and Wilson (1969).

can lead to suboptimal allocations. But the general existence of Pareto optimal outcomes reached via trade when coalitions are unconstrained leaves room for a more optimistic perspective.

A criterion more demanding than Pareto optimality is the correspondence between outcomes generated by Pivot stable vote allocations and the Condorcet winner–the outcome that a majority of voters prefer to any other–if it exists. We analyze such a correspondence, called Condorcet consistency in the social choice literature[3], in our third set of results. In general, Pivot trading is not Condorcet consistent: even when the Condorcet winner exists, trading may lead to a stable outcome that differs from the Condorcet winner. Special cases exist–e.g., if there are only three voters, or two proposals–such that vote trading is guaranteed to deliver the Condorcet winner, but the result does not hold more broadly. The connection between outcomes generated by Pivot stable allocations and the Condorcet winner thus is tenuous: we know that the former always exist while the latter typically does not, and even when the latter exists, vote trading need not deliver it.

The link between vote trading and the Condorcet winner was a central unresolved question in the early literature. Buchanan and Tullock (1962) and Coleman (1966) conjectured that vote trading offers the solution to majority cycles in the absence of a Condorcet winner, a belief we find still expressed in popular writings on voting.[4] Starting with Park (1967), a number of authors studied and rejected the conjecture[5], but the different scenarios and the incompletely specified trading rules make comparisons difficult. Our existence result can be read as partially supporting Buchanan and Tullock's, and Coleman's conjecture. But the connection is weak because the logic in the older arguments seems quite different and, contrary to the implicit claims of all authors cited above, existence of a Condorcet winner in general does not imply that it must be reached by vote trading.

As our description makes clear, the object of our study is the trade of votes for votes within a committee, in the absence of side-payments. Thus the model and the approach are quite different from the rich literature analyzing the trade of votes in exchange for a numeraire, whether vote buying by candidates or lobbyists (Myerson (1993), Groseclose and Snyder (1996), Dal Bo (2007), Dekel et al. (2008, 2009)), or vote markets (Philipson and Snyder (1996), Casella et al. (2012), Xefteris and Ziros (2017)), or auction-like mechanisms (Lalley and Weyl (2016), Goeree and Zhang (2017)).

The lack of side-payments evokes instead the work on alternative voting rules that allow outcomes to reflect intensity of preferences. The literature includes the storable votes mechanism of Casella (2005), qualitative voting (Hortala-Vallve (2012)), and the linking mechanisms proposed in Jackson and Sonnenschein (2007). There are however two major differences. First, in these schemes voters can shift their own votes from one proposal to another, within the limits of a budget constraint, but are not allowed to trade votes with other voters. Second, such mechanisms

[3] See Moulin (1988).

[4] "If logrolling is the norm, then the problem of the cyclical majority vanishes." (Buchanan and Tullock, 1965 ed., p.336.). "When logrolling is allowed, the highest valued outcome is secure without the threat of a cyclical majority." (https://en.wikipedia.org/wiki/Logrolling, accessed June 201, 2018)

[5] See also Bernholz (1973), Ferejohn (1974), Koehler (1975), Schwartz (1975), Kadane (1972), Miller (1977).

are formulated as solutions to Bayesian collective decision problems, where preference intensity is represented by von Neumann-Morgenstsern utility functions. The approach and solution concepts are grounded in non-cooperative game theory and agents maximize expected utility. Neither feature applies to our analysis, where votes can be traded across voters but not across proposals, and preferences representations that maintain ordinal rankings are fully interchangeable.

In terms of solution concepts, this paper is connected to work on dynamics and stability in environments that do not allow side payments. We have in mind the problem of achieving stability in sequential rounds of matching among different agents (Gale and Shapley, 1962; Roth and Sotomayor, 1990; Roth and Vande Vate, 1990), in creating or deleting links in the formation of networks (Jackson and Wolinsky, 1996; Watts, 2001; Jackson and Watts, 2002), or in sequences of barter trades in an exchange economy without money (Feldman, 1973, 1974; Green, 1974). While the substantive issues addressed in those paper are different from vote trading, the modeling of dynamics and stability is similar in spirit to ours. In all of these cases, as in the approach we take in this paper, the problem is studied by combining a definition of stability and a rule specifying the dynamic process leading to stable outcomes.

In what follows, we begin by describing the general framework (Section 2). In Section 3, we discuss the existence of stable vote allocations reachable via trading, and their properties–the Pareto optimality of stable vote allocations and the relationship between stable outcomes and the Condorcet winner. Section 4 summarizes our conclusions and discusses possible directions of future research.

2 The model

Consider a committee $\mathcal{C} = \{1, ..., N\}$ of N (odd) voters who must approve or reject each of K independent binary proposals. The set of proposals is denoted $P = \{1, ..., k, ..., K\}$. Committee members have separable preferences represented by a profile of values, z, where $z_i^k \in \mathbb{R}$ 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. The value of any proposal failing is normalized to 0. We call $x_i \equiv |z_i|$ voter i's *intensity* on proposal k. We specify the profile of cardinal values z because working with such a profile will prove convenient and intuitive, but our analysis relies only on individual ordinal rankings over the 2^K possible outcomes (all possible combinations of passing and failing for each proposal). 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. One can think of votes in our model as if they were physical ballots, each one tagged by proposal. A vote trade is an exchange of ballots, with no enforcement or credibility problem. After trading, a voter may own zero votes over some proposals and several votes over others, but cannot hold negative votes on any issue. We call v_i^k the votes held by voter i over proposal k, $v_i = (v_i^1, ..., v_i^K)$ the profile of votes held by i

over all proposals, and $v = (v_1, ..., v_i, ..., v_N)$ a *vote allocation*, i.e., a profile of vote holdings for all voters and proposals. The initial vote allocation is denoted by $v_0 = (v_{01}, ..., v_{0N})$. We impose no restriction on v_0, beyond $v_{0i}^k \geq 0$ for all i, k and, to avoid ties, $\sum_i v_{0i}^k$ odd for all k. Let $\mathcal{V}$ denote the set of feasible vote allocations: $v \in \mathcal{V} \Longleftrightarrow \sum_i v_i^k = \sum_i v_{0i}^k$ for all k and $v_i^k \geq 0$ for all i, k.[6]

Definition 1 *A **trade** is an ordered pair of vote allocations (v, v'), such that $v, v' \in \mathcal{V}$ and $v \neq v'$.*

That is, the trade (v, v') is a reallocation of votes from v to v'. Voter i's *net trade* from (v, v') is denoted $\delta_i(v, v')$, where $\delta_i^k(v, v') = v_i'^k - v_i^k$.

Given a feasible vote allocation v, when voting takes place on proposal k each voter has a dominant strategy to cast all her votes in favor of the proposal if her proposal's value is positive ($z_i^k > 0$), and against the proposal if her proposal's value is negative ($z_i^k < 0$). We indicate by $\mathbf{P}(v) \subseteq P$ the set of proposals that receive at least $(\sum_i v_{0i}^k + 1)/2$ favorable votes, and therefore pass. We call $\mathbf{P}(v)$ the *outcome* of the vote if voting occurs at allocation v. Finally, we define $u_i(v)$ as the utility of voter i if voting occurs at v: $u_i(v) = \sum_{k \in \mathbf{P}(v)} z_i^k$. Preferences over outcomes are assumed to be strict. That is, $u_i(v) = u_i(v')$ if and only if $\mathbf{P}(v) = \mathbf{P}(v')$.[7]

Our focus is on the existence and properties of vote allocations that hold no incentives for trading. Without loss of generality we will assume that there are no unanimous issues, i.e., there is no issue k such that either $z_i^k > 0$ for all i or $z_i^k < 0$ for all i.[8] Consider any trade (v, v'), and let $\Delta_i^k(v, v') = |\delta_i^k(v, v')|$ denote the absolute change in vote holdings for individual i on proposal k. Denote $\Delta_i(v, v') = \sum_{k=1}^K \Delta_i^k(v, v')$. We define:

Definition 2 *Let $C \subseteq \mathcal{C}$ be a non-empty coalition. The trade (v, v') is a **payoff-improving trade for** C if $i \in C \Leftrightarrow \Delta_i(v, v') > 0$ and $i \in C \Rightarrow u_i(v') > u_i(v)$.*

That is, a trade is called *payoff-improving for C* if only voters in C, and all voters in C, trade, and every voter in C is strictly better off with the outcome that would result from the new vote allocation. Coalition C can have any arbitrary size between 2 and N. We then say:

Definition 3 *A coalition $C \subseteq \mathcal{C}$ **blocks** v if there exists a payoff-improving trade (v, v'), for C. Call (v, v') a **blocking trade**.*

We denote by $B(v)$ the set of all blocking trades at v – i.e. the set of all feasible payoff-improving trades for all possible coalitions.

Definition 4 *A vote allocation $v \in \mathcal{V}$ is **stable** if $B(v) = \varnothing$.*

[6] Note that $\sum_k v_i^k \neq \sum_k v_{0i}^k$ is feasible because we do not restrict trades to be one-to-one.

[7] For any i, strictness is satisfied for all z_i, except for a set of measure zero. Some of the examples considered later in the paper allow voters to have weak preferences. This is done for expositional clarity only, and the examples are easily modified to strict preferences.

[8] Exchanging votes on a unanimous issue can never change the outcome.

Our definition of stability thus coincides with the core: a vote allocation $v \in \mathcal{V}$ is stable if it belongs in the core. Note that for any N, K, and z the core is not empty: a feasible allocation of votes where a single voter i holds a majority of votes on every issue is always in the core and thus is trivially 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. Hence:

Proposition 1 *A stable vote allocation v exists for all z, N, and K.*

2.1 Dynamic adjustment: Pivot algorithms.

Stable vote allocations exist, but are they reachable through sequential decentralized exchange? To answer the question, we need to specify the dynamic process through which trades take place.

We posit a dynamic process characterized by sequences of trades yielding myopic strict gains to all coalition members:

Definition 5 *A **Pivot algorithm** is any mechanism generating a sequence of trades as follows: Start from the initial vote allocation v_0. If there is no blocking trade, stop. If there is one such trade, execute it. If there are multiple such trades, execute one according to a choice rule R. Continue in this fashion until no further blocking trade exists.*

The definition describes a family of algorithms, and individual algorithms differ in the specification of the choice rule R that is applied when multiple blocking trades are possible. For example, R may select each possible trade with equal probability; or give priority to trades with higher total gains or involving fewer, or more numerous, or specific voters. Rule R can depend on the current allocation or history of votes, and can be stochastic. Formally, R specifies a probability distribution over $B(v)$, for each vote allocation v such that $B(v) \neq \varnothing$. For any $B(v) \neq \varnothing$ and for any $(v, v') \in B(v)$, we denote by $R(v, v') \geq 0$ the probability that (v, v') is selected at v, with $\sum_{v' \in B(v)} R(v, v') = 1$. Note that R selects a trade, hence both a coalition and a specific exchange of votes for that coalition, among all possible coalitions and vote exchanges that are strictly payoff-improving for the voters involved in the trade.

Payoff improving trades are not restricted to two proposals only, nor to exchanging one vote for one vote: a voter can trade her vote or bundles of votes on one or more issues, in exchange for other voters' vote or votes on one or more issues, or in fact in exchange for no other votes. The only restriction we are imposing is that the trades be strictly payoff-improving for all traders. If a trade is payoff improving, it is a legitimate trade under the Pivot algorithms.

The name *Pivot algorithm* comes from an observation due to Riker and Brams (1973): if a trade is strictly payoff improving, it must alter the outcome of the vote; hence it must involve pivotal votes. In the broad definition we are using here, not all traded votes need to be pivotal: as long as some are, and the outcome is modified by the trade in a direction that benefits all members of the trading coalition, redundant votes may be traded too. Redundant votes are votes whose trade does not affect the outcome: votes traded between voters on the same side of an issue, or votes traded between voters on opposite sides, but not sufficient to change which side holds a majority. Their trade is irrelevant to myopic payoffs but can affect the path of future trades by altering the blocking possibilities of different future coalitions.

3 Pivot-stable vote allocations: Existence and Properties

Do Pivot algorithms converge to stable vote allocations? Stated differently, do sequential myopic trades converge to the core? The question is not trivial because any Pivot trade changes default outcomes and alters the existing set of blocking trades, potentially leading to new Pivot trades, in a sequence that in principle could result in a perennial cycle.

3.1 Existence

We define:

Definition 6 *An allocation of votes v is **Pivot-stable** if it is stable and reachable from v_0 through a Pivot algorithm in a finite number of steps.*

The following result establishes the general existence of Pivot-stable vote allocations:

Theorem 1 *A Pivot-stable allocation of votes exists for all v_0, K, N, and z.*

Before presenting a formal proof of the theorem, it is useful to first explain the intuition with an example.

Example 1. Consider the value matrix in Table 1: rows represent proposals, columns represent voters, and the entry in each cell is z_i^k, the value attached by voter i to proposal k passing. (Recall that the value of a failed proposal is normalized to zero for all voters.)

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

Table 1: *Value matrix for Example 1.*

Suppose $v_0 = \{\mathbf{1}, \mathbf{1}, ..\mathbf{1}\}$. At v_0, proposals A and B both pass with a vote of $3-2$: $u_1(v_0) = u_2(v_0) = 1$, $u_3(v_0) = u_4(v_0) = -1$, $u_5(v_0) = 3$. Allocation v_0 is not stable: it can be blocked by voters 3 and 4. Voter 3 gives a B vote to 4, in exchange for an A vote, reaching a new vote allocation $v_1 = \{\{1,1\},\{1,1\},\{2,0\},\{0,2\},\{1,1\}\}$. At v_1 both proposals fail, and $u_3(v_1) = u_4(v_1) = 0$, a strict payoff improvement for voters 3 and 4. Voters 3 and 4 have shifted votes away from a lower value proposal each was, pre-trade, winning towards a higher value proposal each was losing. The difference in values is the key to the payoff-improving nature of the trade. Vote allocation v_1 is not stable either. Voters 1 and 2 can block it: voter 1 can trade her B vote to 2 in exchange for an A vote, reaching allocation $v_2 = \{\{2,0\},\{0,2\},\{2,0\},\{0,2\},\{1,1\}\}$, such that both proposals pass, and $u_1(v_2) = u_2(v_2) = 1$, a strict payoff improvement for 1 and 2 over allocation v_1. Again, the logic of the trade is a shift in votes from low-value proposals the voters were winning pre-trade to higher value proposals the voters were losing. Allocation v_2 is stable.

Over the sequence of trades, the voters' myopic payoffs have moved non-monotonically, falling and then rising for voters 1, 2, and 5, rising and then falling for voters 3 and 4. The changes in payoffs reflect both the direct gains from the trades the voters themselves have executed and the externalities caused by others' trades. The number of votes held on each proposal, on the other hand, is affected only by the trades a voter participates in. At each step of the process, we can construct an index of the total potential value of each voter's vote holdings, independently of the voting outcome. Specifically, let this index be defined as the intensity-weighted sum of i's votes–here $x_i^A v_i^A + x_i^B v_i^B$, and call it i's *score* at v.[9]

Voter i's score does not change when i does not trade (by construction) and, at least in this example, increases whenever i does trade: after the first trade, it rises from 3 to 4 for voters 3 and 4 (the two voters who trade); after the second trade, it rises, again from 3 to 4, for voters 1 and 2. The increases reflect the logic of the payoff-improving trades. But note that for each voter, the index has a finite ceiling. In this example, the ceiling is $5(x_i^A + x_i^B)$, where 5 is the total number of existing votes on each issue. Thus, if each voter's score can only move monotonically upwards, trade must end in finite time.

What complicates the proof of Theorem 1 is that, unfortunately, the simple monotonicity of the example does not extend to the general case. If multiple votes are given away on the same proposal, if trades involve more than two voters, if some of the votes traded are redundant, in all of these cases traders in payoff-improving trades may see their scores decline. Consider Example 2:

Example 2. The value matrix is reported in Table 2. As before, rows represent proposals, columns represent voters, and the entry in each cell is z_i^k, the value attached by voter i to proposal k passing.

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

Table 2: *An example where the only existing blocking trade causes a decline in score for voter 1. Vote allocations are { {1,2},{1,0},{1,1} }*

[9] One can see that separability is essential to the construction, as the score function is only well-defined with separable preferences.

At v, voter 1 has one vote on A and two votes on B; voter 2 has one vote on A and zero votes on B, and voter 3 has one vote on each proposal. Without trade, A fails 2-1 and B fails 3-0. But v is not stable: voters 1 and 2 are a blocking coalition. Voter 1 trades both of her B votes to voter 2 in exchange for voter 2's A vote; in the resulting allocation v' both A and B pass 2-1, and $u_1(v') = u_2(v') = 1$, a strict improvement for both voters over $u_1(v) = u_2(v) = 0$. Nevertheless, voter 1's score falls from 13 to 10. There is no alternative trade that benefits all members of a trading coalition, and thus there is no trade such that voter 1's score does not fall.

It turns out, however, that the main intuition is robust. The simple fact that all voters taking part in a trade must strictly gain from the trade is sufficient to guarantee that there always exists a path of trades such that any voter's score can decline at most a finite number of times. But then, since the score is bounded, trade must end in finite time. The proof of Theorem 1 defines a general algorithm for constructing such a path for arbitrary environments.

The proof proceeds in two steps. We begin by side-stepping the complication illustrated in Example 2: Lemma 1 shows that if every blocking trade changes outcomes only on proposals that win or lose by exactly one vote, then for every blocking coalition C and for every $i \in C$ there always exists a blocking trade for C that is score-improving for i. The second part of the proof then expands the environment to arbitrary v, allowing for blocking trades in which multiple votes may be traded away.

Before describing the proof, two additional definitions are useful. First, the index used in Example 1 should be defined formally:

Definition 7 *Consider voter i and a vote allocation v. Voter i's **score** at v is given by:*

$$\sigma_i(v) = \sum_{k=1}^{K} x_i^k v_i^k.$$

Second, Lemma 1 applies to blocking trades on proposals decided by a single vote. This too should be made precise.

Definition 8 *Call N_+^k the set of voters in favor of proposal k, and N_-^k the set of voters against proposal k. We say that at v a proposal is **decided by minimal majority** if $\left|\sum_{i \in N_+^k} v_i^k - \sum_{i \in N_-^k} v_i^k\right| = 1$.*

Lemma 1 *Suppose that at v every blocking trade (v, v') changes outcomes only on proposals that are decided by minimal majority at v. Then for any C that blocks v and for any $i \in C$, there exists a blocking trade, (v, v'), such that $\sigma_i(v') > \sigma_i(v)$.*

Proof. The proof is constructive. If v is stable, there are no blocking trades. Suppose then that v is not stable and there is at least one blocking coalition; if there is more than one, select any blocking coalition C. Because C is a blocking coalition at v, there must exist at least one set of (two or more) proposals whose resolution is modified by a feasible payoff improving trade within C. If multiple sets of such proposals exist, select one. Call it $\widetilde{P}$. Consider any voter $i \in C$. Define $\widetilde{P}^{w(i)} = \{P \in \widetilde{P} \mid i$ is on the winning side for $P \in \widetilde{P}$ post-trade and is on the losing side pre-trade$\}$ and $\widetilde{P}^{l(i)} = \{P \in \widetilde{P} \mid i$ is on the losing side for $P \in \widetilde{P}$ post-trade and is on the winning side pre-trade$\}$, and observe that $\widetilde{P} = \widetilde{P}^{w(i)} \cup \widetilde{P}^{l(i)}$, since, by selection of $\widetilde{P}$, trade changes the resolution of all proposals in $\widetilde{P}$. Because $i \in C$, it must be the case that i strictly gains from the trade overall. Hence, even though the two sets, $\widetilde{P}^{w(i)}$and $\widetilde{P}^{l(i)}$, may have different cardinality, by definition of improving trade, $\sum_{k\in\widetilde{P}^{w(i)}} x_i^k < \sum_{k\in\widetilde{P}^{l(i)}} x_i^k$. Because all $P \in \widetilde{P}$ are decided by minimal majority at v, one can construct a blocking trade by reallocating a single vote within C on each $P \in \widetilde{P}$, and leaving unchanged all vote holdings on the other proposals. Specifically, construct a trade such that i receives one extra vote on all $P \in \widetilde{P}^{w(i)}$, and gives away one vote on any $P^k \in \widetilde{P}^{l(i)}$such that $v_i^k > 0$. But then

$$
\begin{aligned}
\sigma_i(v') - \sigma_i(v) &= \left(\sum_{k\in\widetilde{P}^{w(i)}} x_i^k v_i'^k + \sum_{k\in\widetilde{P}^{l(i)}} x_i^k v_i'^k\right) - \left(\sum_{k\in\widetilde{P}^{w(i)}} x_i^k v_i^k + \sum_{k\in\widetilde{P}^{l(i)}} x_i^k v_i^k\right) \\
&\geq \sum_{k\in\widetilde{P}^{w(i)}} x_i^k - \sum_{k\in\widetilde{P}^{l(i)}} x_i^k \\
&> 0.
\end{aligned}
$$

The score of voter i has increased. ■

An observation about this construction is key to understanding what follows. Voter i, designated as the recipient of a vote for each proposal in $\widetilde{P}^{w(i)}$ and, wherever possible, as the source of the traded votes for proposals in $\widetilde{P}^{l(i)}$, is chosen arbitrarily and can be any member of the blocking coalition. The trade is constructed to guarantee that i's score increases: for any arbitrary $i \in C$, there exists a trade with such property.

Proof of Theorem 1. We construct an algorithm such that, starting at any initial vote allocation v_0, for any K, N, and z, there exists a finite sequence of blocking trades ending in a stable vote allocation v^*.

At any step of the sequence with vote allocation v, denote by $\widehat{P}(v)$ the set of proposals that are <u>not</u> decided by a minimal majority at v, with $|\widehat{P}(v)| \leq K$. There are three cases to consider. Case 0: there exists no blocking trade. Hence v is stable, and the theorem holds. Case 1: there exists a blocking trade which changes the outcomes on some proposals that are not decided by minimal majority at v. Case 2: all blocking trades at v change only proposals that are decided by a minimal majority at v.

If we are in Case 1 at v, there exists at least one blocking trade that includes exchanging pivotal votes on a non-empty subset of $\widehat{P}(v)$. If there are multiple such trades, select one, and call $\widehat{\widehat{P}}$ the set of non-minimal majority proposals at v whose resolution is modified by the trade. Any outcome achieved by a blocking trade involving $\widehat{\widehat{P}}$ can always be replicated by a blocking trade,

(v, v'), constructed so that at v' all proposals in $\widehat{\widehat{P}}$ are decided by a minimal majority. Execute one such trade. This reduces the number of proposals that are not decided by a minimal majority by $|\widehat{\widehat{P}}| > 0$. Thus $|\widehat{P}(v')| < |\widehat{P}(v)|$. At v', again we can be in Case 0, Case 1, or Case 2.

If we are in Case 2 at v, then there exists a blocking coalition and a blocking trade for that coalition, (v, v'), that only changes proposals decided by minimal majority at v. If there are multiple such coalitions, select one, and call it C. Assign to each voter an index $i \in \{1, ..., N\}$, and define i^*_C to be the unique voter in C with the property that $i^*_C \leq i$ for all $i \in C$–that is, i^*_C is the voter in C with the lowest index. By Lemma 1 we can find a blocking trade for C such that $\sigma_{i^*_C}(v') > \sigma_{i^*_C}(v)$, and such that the proposals involved in the trade continue to be decided by minimal majority at v'. Execute that trade. At v', again we can be in Case 0, Case 1, or Case 2.

At any future step and vote allocation v proceed as above. The algorithm defines a sequence of blocking trades, or ends trade if no blocking trade exists. We claim that this algorithm must end after a finite number of trades.

The logic is as follows. First, because $|\widehat{P}(v)| \leq K < \infty$ and $|\widehat{P}(v')| < |\widehat{P}(v)|$ we can only be in Case 1 a finite number of times in the sequence. Thus we only have to ensure that we can be in Case 2 only a finite number of times. Consider voter 1. For voter 1, we know that whenever we are in Case 2 at step t, $\sigma_1(v') > \sigma_1(v)$ if $1 \in C$, because $1 = i^*_C$, and $\sigma_1(v') = \sigma_1(v)$ if $1 \notin C$. Because 1's score is a bounded function of v, this implies that 1 can be in at most a finite number of Case 2 blocking trades. From above, we also know that 1 can be party to at most a finite number of Case 1 blocking trades. Hence there is a finite number of steps in the sequence that have a blocking trade with a coalition that includes voter 1.

Next consider voter 2. For voter 2 we know that $\sigma_2(v') > \sigma_2(v)$ whenever we are in Case 2 and $1 \notin C$ but $2 \in C$, because $2 = i^*_C$. At any step of Case 2 where $\{1, 2\} \subseteq C$, 2's score may possibly decrease because $1 = i^*_C$, but this can happen only a finite number of times, because 1 can only be in a finite number of blocking trades in the sequence. At any step of Case 2 where $2 \notin C$, 2's vote holdings are unchanged, so 2's score is unchanged. Because 2's score is bounded above, this implies that 2 can be in at most a finite number of Case 2 blocking trades. And, from above, we also know that 2 can be involved in at most a finite number of Case 1 blocking trades. Hence there is a finite number of steps in the sequence that have a blocking trade involving voter 2. Extending the logic of this argument to voters with indices $i > 2$, it follows that every voter can be in at most a finite number of blocking trades in the sequence. Because there is a finite number of voters, each of whom can be involved in only a finite number of blocking trades in the sequence, the sequence can only have a finite number of steps, and must end at a stable vote allocation. Hence the set of Pivot stable allocations is non-empty. ■

The result holds broadly. The only condition we impose is that all members of a trading coalition must strictly benefit (myopically) from the trade. We do not restrict the size of the coalition or the number of proposals affected by vote trades; we do not require that trades be one-to-one or limited to pivotal votes. And yet we find that there is always–for any number of voters and proposals, for any profile of separable preferences and any initial vote allocation–a path of payoff-improving trades that leads to a stable vote allocation. Note that because the theorem holds for any arbitrary selection of coalition C, it holds, a fortiori, if we constrain allowable coalitions–for example if we allow only pairwise trades or impose some cohesion requirement on C. Any such constraint will reduce the set of unstable vote allocations and strengthen the case for stability.

The theorem does not say that every trading path must converge to stability; rather, it says that there always exists a trading path for which this is true. However, the existence of one such path for any arbitrary starting allocation v_0 allows us to identify a broad class of selection rules R for which convergence to stability is guaranteed to occur in finite time. Call R_r the family of all rules R such that, for all $v \in \mathcal{V}$ and for all $(v, v') \in B(v)$, $R(v, v') > 0$. That is, R_r is the family of all stochastic selection rules R that put positive probability on any existing blocking trade. We can then state:[10]

Corollary 1 *If $R \in R_r$, then for all v_0, K, N, and z, a Pivot-stable allocation of votes is reached with probability 1 in finite time.*

Proof. For any vote allocation v, if v is stable, the result holds trivially; if not, denote by $L(v)$ the length of the shortest sequence of blocking trades, starting at v and ending at some stable vote allocation $v^*_{L(v)}$. Let $\overline{L} = \max_{v \in \mathcal{V}}\{L(v)\}$, which we know exists because $\mathcal{V}$ is a finite set and a stable vote allocation v^* exists. Let $r = \min\{R(v, v') | B(v) \neq \varnothing \text{ and } (v, v') \in B(v)\} > 0$, and let $\pi = r^{\overline{L}}$(where $\overline{L}$ is a power). Suppose the initial vote allocation is v_0. Then the probability that a stable allocation is reached in a sequence of $\overline{L}$ or fewer trades from v_0 is greater than or equal to π. Similarly the probability that a stable allocation is reached in a sequence of $m\overline{L}$ or fewer trades from v_0 is greater than or equal to $\sum_{j=1}^{m}(1-\pi)^{j-1}\pi = \pi\frac{1-(1-\pi)^m}{1-(1-\pi)} = 1-(1-\pi)^m \to 1$ as $m \to \infty$. ■

No additional condition is required. As long as all trades have some chance of being selected, the result holds: convergence to a stable allocation will occur in a finite number of steps.

3.2 Pairwise Trading

Theorem 1 and its corollary tell us that vote trading will lead to stability for a large class of selection rules, in arbitrary environments. But can we identify conditions under which convergence is guaranteed for all selection rules? Riker and Brams (1973) proposed a trading rule not unlike our Pivot algorithms–payoff-improving, myopic, enforceable trades–and conjectured that convergence to stability required limiting trades to be pairwise. Theorem 1 shows that the conjecture is incorrect. And yet we show in this section that restricting trade to be pairwise can lead to a stronger result. When complemented with one intuitive additional condition, pairwise trading leads to stability along *all* trading paths.

[10]The intuition is similar to a well-known result in the matching literature: in marriage markets, random matching algorithms will eventually lead to a stable match (Roth and Vande Vate 1990). Note however that our environment is quite different, primarily because payoffs depend on the entire profile of vote allocations.

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

Table 3: *Pairwise Pivot trades need not converge if redundant trades are possible. An example.*

The additional condition excludes the trade of redundant votes–the gratuitous exchange of votes that have no effect. Once again, it is required to preserve the monotonicity of the score function along the path of trade. Consider the following example:

Example 3. Suppose $v_0 = \{\mathbf{1}, ..\mathbf{1}\}$. There are four proposals and five voters, and the value matrix is shown in Table 3:

At v_0, proposals A, B, and D pass; proposal C fails. All proposals pass or fail by minimal majority. Consider the following sequence of pairwise Pivot trades. At v_0, voter 1 gives one A vote and one D vote to 2, in exchange for one B vote and one C vote. The trade is strictly payoff-improving for both traders because it alters the majority direction on A and B; it does not alter the voting tally on C and D, on which 1 and 2 agree. At v_1, proposal D passes and all others fail. Voters 2 and 3 can block v_1: voter 2 gives one A vote and one D vote to 3, in exchange for one C and one B vote. The trade alters the majority on A and C, and is payoff-improving for both voters; it does not affect the resolution of D and B, on which the two voters agree. At v_2, A, C, and D pass, and B fails. But v_2 is not stable: voters 3 and 4 can trade and raise their myopic payoff. Voter 3 gives one D vote and one A vote to 4, in exchange for one B and one C vote. The majority changes on B and D, but not on A and C, on which the two voters agree. At v_3, A, B, and C pass, and D fails. But voters 1 and 4 can block v_3: voter 1 gives one C and one B vote to 4, in exchange for one D and one A vote. The trade is strictly payoff-improving because it alters the majority on C and D in the direction both traders prefer; it does not alter the majority on A and B, on which the two traders agree. This last trade, however, has brought the vote allocation back to v_0. The sequence of trades can then be repeated into a never ending cycle.

In Example 3, all Pivot trades are pairwise and all proposals, at any step on the path of trade, are decided by minimal majority. Yet, it is readily verified that traders' scores at times decrease, and vote allocations cycle. The problem comes from vote trades on proposals on which the traders agree. These redundant trades have no effect on payoffs, and thus a voter can trade away a vote on a high value proposal for a vote on a lower value proposal: the trade has no effect, but the voter's score declines. The declines in score make cycles possible.

To guarantee convergence to a stable vote allocation, we need to rule out redundant trades. What this means exactly is formalized in the following two definitions.

Definition 9 (v, v'') *is a reduction of* (v, v') *if* $P(v'') = P(v')$, $\Delta_i^k(v, v') = 0 \Rightarrow \Delta_i^k(v, v'') = 0$ *for*

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

Table 4: *Minimal Pivot trades need not converge if coalitional trades are possible. An example.*

all i, k, and for all i, k, $\Delta_i^k(v, v') \geq 0 \Rightarrow \Delta_i^k(v, v') \geq \Delta_i^k(v, v'')$, with $\Delta_i^k(v, v') > \Delta_i^k(v, v'')$ for some i, k.

Definition 10 *Consider a blocking trade (v, v'). We say that (v, v') is a minimal blocking trade if there does not exist a reduction of (v, v').*

Loosely speaking, minimality rules out two kinds of trades: if trade (v, v') does not change the outcome of proposal k, then no votes are traded on k; and if trade (v, v') does change the outcome of proposal k, then k is decided by minimal majority at v'. It is straightforward to show that if v is not a stable allocation, then the set of minimal blocking trades is non-empty.

We can then state:

Theorem 2 *If trades are restricted to be pairwise and minimal, then a Pivot-stable allocation of votes exists for all v_0, K, N, z, and R.*

As in the case of Theorem 1, the proof builds on the score function. It shows that, in the streamlined environment of Theorem 2, traders' scores can decrease only if trade occurs on non-minimal majority proposals. But by minimality, any such trade must bring the proposals to minimal majority, and thus the number of trades on which scores can fall must be finite. Because the score function is bounded and the number of voters is finite, it then follows that the number of trades must be finite and bounded as well. And this must be true on any path of trade determined by any choice rule R. A formal proof is in the Appendix.

Example 3 shows that, without minimality, pairwise trade is not sufficient to guarantee convergence to stability for all selection rules. But it is also the case that, without the restriction to pairwise trading, minimality is not sufficient either. Consider the following example:

Example 4. Table 4 reports the value matrix. The initial vote allocation is $v_0 = \{\mathbf{1}, \mathbf{1}, .., \mathbf{1}\}$.

At v_0, all proposals pass by minimal majority, and $u_i(v_0) = -1$ for $i = \{1, 2, 3, 4\}$. Consider a coalition composed of such voters, and the following coalition trade: voter 1 gives her A vote to voter 2, in exchange for 2's B vote; voter 3 gives her C vote to voter 4, in exchange for 4's D vote. At v_1, all proposals fail and $u_i(v_1) = 0$ for all coalition members. For all, the trade is strictly improving. The vote allocation v_1 is not Pivot stable: voters 1 and 2 can block v_1 by trading back their respective votes on A and B, reaching outcome $\mathbf{P}(v_2) = AB$, and enjoying a strictly positive increase in payoffs: $u_j(v_2) = 1$ for $j = \{1, 2\}$. At v_2, $u_s(v_2) = -2$ 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: $\mathbf{P}(v_3) = ABCD$, and $u_s(v_3) = -1$ for $s = \{3, 4\}$. 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 that selects the blocking coalitions in the order described, no Pivot stable allocation of votes can be reached.

We can rephrase the observation in terms of each voter's score: although all trades are strictly payoff-improving, the first coalition trade, from v_0 to v_1, lowers the score of all traders involved from 5 to 4. The two successive trades raise the traders' scores back to 5, one pair at a time, but the initial decline makes a cycle possible.

Note that all trades in Example 4 are minimal. And yet, a decline in scores can accompany a profitable trade because of the trading externalities present within the coalition: a coalition member can engage in a vote exchange that by itself would not be profitable and that causes a decline in score because she benefits from the other members' trades. When trade is pairwise and minimal, all trades must be advantageous to all active traders, and this cause of possible cycles is excluded.

We conclude this section with one final remark on the technique used to prove and illustrate our results. We have relied repeatedly on the score function because it makes transparent the source of the gain from blocking trades and the built-in ceiling in such possible gains and trades. The score function is a cardinal measure of the potential value of voters' vote holdings, but it is important to stress that the reliance on a cardinal measure is for convenience only. The logic is fully ordinal: changing all intensities x_i^k in any arbitrary fashion that preserves all ordinal rankings has no impact on any of our results.

3.3 Properties of Pivot-stable outcomes

Following Theorem 1, a Pivot-stable vote allocation always exists. When trade comes to an end, the outcome of the vote is realized. Do outcomes reached via vote trading possess desirable welfare properties?

We define:

Definition 11 *An outcome* $\mathbf{P}(v)$ *is a* ***Pivot-stable outcome*** *if* v *is a Pivot-stable vote allocation.*

For any fixed K, N, and z, we denote $\mathcal{V}^*$ the set of all Pivot-stable vote allocations, and $\mathcal{P}(\mathcal{V}^*)$ the set of all stable outcomes reachable with positive probability through a Pivot algorithm. If $\mathcal{P}(\mathcal{V}^*)$ is a singleton, we use the notation $\mathbf{P}(\mathcal{V}^*)$ to denote the unique element of $\mathcal{P}(\mathcal{V}^*)$.[11]

We find:

Theorem 3 *All Pivot-stable outcomes must be in the Pareto set, for all* v_0*,* K*,* N*, and* z*.*

[11] Note that uniqueness of $\mathcal{P}(\mathcal{V}^*)$ does not imply that $\mathcal{V}^*$ is a singleton. There can be multiple Pivot-stable vote allocations, all leading to the same outcome.

Proof. We know from Theorem 1 that a Pivot-stable outcome exists. Regardless of the history of previous trades, if the outcome is Pareto dominated, then the coalition of the whole can always reach a Pareto superior outcome and has a profitable deviation. But then the allocation corresponding to the Pareto-dominated outcome cannot be Pivot-stable. ■

If no restriction on coalition formation is imposed, all Pivot-stable outcomes must be Pareto optimal. Coupled with Theorem 1, Theorem 3 teaches us that vote trading can always reach a Pareto optimal outcome (contrary to earlier conjectures). But note that the result relies on being able to form the coalition of the whole. Thus the result holds for any coalitional restriction that does not interfere with the coalition of the whole, but does not hold if such coalition cannot form. Riker and Brams' (1973) "paradox of vote trading" is a well-known example where pairwise trades only are possible, and the outcome they identify (which would be Pivot-stable if only pairwise trades were allowed) is not Pareto optimal.[12]

A second property generally viewed as desirable in voting environments is the ability to reach the Condorcet winner: the outcome that is preferred by a majority of voters to every other outcome. The Condorcet winner need not exist, and a voting system is said to satisfy Condorcet consistency if it uniquely selects the Condorcet winner whenever it does exist. Is Pivot stability Condorcet consistent?[13]

Because the Condorcet winner implicitly assumes unweighted voting, in the remainder of this section we restrict the analysis to environments where $v_0 = \{1, ., 1\}$**.** The main result is negative: vote trading may lead to stable outcomes that differ from the Condorcet winner.

Proposition 2 *If $K > 2$ and $N > 3$, there exist z such that $\mathbf{P}$ is the Condorcet winner but there exists $\mathbf{P}' \neq \mathbf{P}$ such that $\mathbf{P}' \in \mathcal{P}(\mathcal{V}^*)$.*

Proof. Consider the following environment with $v_0 = \{\mathbf{1}, .., \mathbf{1}\}$, $K = 3$ and $N = 5$:

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

Table 5: *Preference profile such that a Pivot-stable outcome is not the Condorcet winner.*

For the preference profile in Table 5, $\mathbf{P}(v_0) = ABC$ is the Condorcet winner. Consider the following set of trades. At v_0, voter 2 gives a B vote to 3, in exchange for a A vote; at the new vote allocation $\mathbf{P}(v_1) = C$. However v_1 is not stable: it can be blocked by voters 4 and 5, trading votes on A and B so that $\mathbf{P}(v_2) = ABC$. But v_2 is again not stable: it can be blocked by 1 and 3,

[12] It is also easy to construct cases in which a stable outcome reached via pairwise trade is in fact Pareto optimal. The general point is that with pairwise trade Pareto optimality is not guaranteed.

[13] Utilitarian welfare criteria are not appropriate here because they depend on cardinal preferences, and thus can vary for fixed ordinal rankings.

trading votes on B and C so that $\mathbf{P}(v_3) = A$. Allocation v_3 is stable, and thus A is a Pivot-stable outcome along this path of trade. To see that v_3 is stable, notice that no B votes can be traded because voter 3 has a majority of B votes and ranks winning B higher than winning A and C. Thus at v_3 the proposals on which trade can possibly occur are only two, A and C, and the possible outcomes are A (the outcome at v_3) or C. Voter 3 cannot trade votes away because she has 0 votes on both A and C. As for the other voters, either they do not want to trade because they prefer A to C (voters 1, 4 and 5) or only hold losing votes and cannot trade (voter 2). Finally, note that although a majority prefers $ABC = \mathbf{P}(v_0)$ to $A = \mathbf{P}(v_3)$, the no-trade allocation v_0 is not stable.

The example can be generalized to an arbitrary number of voters and proposals. Maintaining first $K = 3$, we can add to the example any even number of voters such that at v_3 half of them win on all proposals (i.e. prefer A to pass, and B and C to fail) and half of them lose on all proposals (i.e. prefer A to fail, and B and C to pass). Adding such voters cannot induce any further trade at v_3. As long as their preferences are such that both types of voters rank ABC above both outcome B and outcome C, ABC remains the Condorcet winner. And yet $\mathbf{P}(v_3) = A$ remains Pivot stable. We can then extend the example to arbitrary $K > 3$ by adding proposals such that for each additional proposal, k', voters $i = 1, ..., N-1$ are all in favor of k' passing, and furthermore $z_i^{k'} > x_i^A + x_i^B + x_i^C > 0$ for $i = 1, ..., N-1$. This guarantees that no trade involving these additional proposals will take place, and $\mathbf{P}(v_3) = A$ remains Pivot-stable. ■

Pivot-stability not only fails to satisfy Condorcet consistency; by immediate extension, Proposition 2 implies that Pivot stability is inconsistent with any solution concept that is itself Condorcet consistent.- i.e., that uniquely selects the Condorcet winner when it exists.

The negative result in Proposition 2 does not extend to the special cases of $K = 2$ or $N = 3$. The two propositions below make this point. They are stated separately because the two results stem from very different logic.

Proposition 3 *If $N = 3$, then for all K and z, if the Condorcet winner exists, $\mathcal{P}(\mathcal{V}^*)$ is a singleton, and is the Condorcet winner.*

Proof. See Appendix. ■

Proposition 4 *If $K = 2$, then for all N and z, $\mathcal{P}(\mathcal{V}^*)$ is a singleton and is the Condorcet winner, if the Condorcet winner exists. If $\mathbf{P}(\mathcal{V}^*) \neq \mathbf{P}(v_0)$, a majority prefers $\mathbf{P}(\mathcal{V}^*)$ to $\mathbf{P}(v_0)$.*

Proof. See Appendix. ■

With $N = 3$, the result follows immediately. From Park (1967) and Kadane (1972), we know that the Condorcet winner, if it exists, must coincide with the no trade outcome. If $N = 3$, a pair of voters constitutes a majority, and thus if v_0 delivers the Condorcet winner it cannot be blocked. But Proposition 4 does not follow as directly, because trade is indeed possible. Rather, its proof highlights that in the case of two proposals, trading via the Pivot algorithm can reach only two possible outcomes - the no trade outcome and its complement. This effectively partitions all voters into two groups, with opposite preferences between the no trade outcome and its complement. Differences in preferences ranking over other outcomes within each of these two groups are irrelevant because such outcomes are unreachable. Over reachable outcomes, preferences within each group are perfectly aligned. The scenario thus effectively reduces to a contest between two alternatives, the no trade outcome and its complement, whose resolution is fully determined by which side holds more votes. In contrast, when there are more than two proposals all 2^K possible outcomes are reachable in principle, and it is not possible to partition the voters into two groups with opposite preferences over exactly two reachable outcomes. Hence the logic of the proof of Proposition 4 breaks down for $K > 2$.

4 Conclusions

This paper proposes a general theoretical framework for studying vote trading in committees. It starts from two essential features: (1) a notion of stability: a stable vote allocation is such that no strict payoff improving vote trade exists; and (2) a class of vote trading algorithms–Pivot algorithms–that define dynamic paths from initially unstable vote allocations to stable vote allocations. The model has three key assumptions. First, proposals are binary and preferences are separable across proposals: a voters' preferred resolution of proposal A does not depend on the resolution of proposal B. Second, voting takes place proposal by proposal: voters can trade votes simultaneously on multiple proposals without constraint, but each vote is specialized by proposal. Finally, as in the canonical model of economic exchange, each vote trade is a transfer of a property right among the trading parties: trades cannot be reversed unilaterally, and votes can be re-traded.

The basic framework delivers answers to core questions of existence, Pareto optimality, and Condorcet consistency. The central finding of the paper is a general existence result: there always exists a sequence of payoff-improving trades that leads to a stable vote allocation in finite time, from any initial distribution of votes, for any number of voters and proposals, for any separable preferences, and for any conditions on feasible trading coalitions. Furthermore, if all blocking trades are selected with positive probability, then trading is guaranteed to converge to a stable vote allocation with probability one. In the absence of restrictions on feasible trading coalitions, outcomes corresponding to stable vote allocations must be in the Pareto set, but in general there is no guarantee that trading will result in the Condorcet winner when it exists.

There are a number of interesting directions to pursue, using this framework as a starting point. First, one could extend the myopic trading algorithms to allow for farsighted behavior where voters correctly anticipate the consequences of a trade they engage in, trade which can trigger future trades by other voters. Some initial results are reported in Casella and Palfrey 2018.

Second, the framework can be generalized to allow for voting rules other than simple majority rule, such as qualified majority rule or the existence of veto players. The basic formal concepts such as the definitions of blocking coalitions, blocking trades, stability, Pivot algorithms, and so forth would remain unchanged, although different voting rules would imply different blocking trades.

A third, related extension would consider different restrictions on the vote trading process. Restrictions could be embodied in the contract through which votes are exchanged, for example limiting the extent of re-trading to which a vote is subject. Whether such restrictions would favor or hamper convergence to stability is an open question. Alternatively, restrictions could be imposed on the coalitions that can organize vote trades. We have studied explicitly two possibilities only: unlimited coalitional trading (i.e., no restriction at all on the coalitions that can organize a trade), and, for some additional results, pairwise trading (i.e., any coalition of exactly two voters). But in some committees or legislatures, norms or party ties may limit which coalitions can form. Restricting coalitions will not affect the main existence theorem because restrictions makes blocking more difficult, but could affect the properties of Pivot stable outcomes.

A fourth possible direction concerns agenda setting and agenda manipulation. In the model studied in this paper, the set of binary issues is exogenously given, but **i**n practice, the proposals up for vote are typically the outcome of an agenda formation process. One can imagine different ways to introduce such a process into the model. In one such approach, an agenda setter or committee chair may have the power to bundle proposals. Agenda setting in the form of bundling introduces a different perspective on modeling logrolling in committees. Considering the agenda formation process would move the analysis in the direction of bargaining models of legislative decision making (Baron and Ferejohn (1989)) suggesting a non-cooperative game approach, in contrast to the stability approach pursued here.

Finally, a different but important question is how to incorporate uncertainty in the model. Our framework has no formal inclusion of uncertainty. In a companion paper (Casella and Palfrey, 2016), we report findings from an experiment that reproduces the framework studied here, but where trades are proposed and executed by the ;voters in the experiment, as opposed to being ruled by an algorithm. We find some hoarding of votes on high value proposals, perhaps as a hedge against adverse vote trading by others. This suggests a sensitivity to the strategic uncertainty they face: it is difficult to predict future vote trades that might be triggered by a current vote trade. More traditional modeling of uncertainty using a Bayesian game approach could be explored, incorporating private information either about one's own preferences or about an unknown state of the world that affects everyone's welfare, as in Condorcet jury models.

Appendix. Proofs

Theorem 2. *If trades are restricted to be pairwise and minimal, then a Pivot-stable allocation of votes exists for all* v_0*,* K*,* N*,* z*, and* R*.*

Proof. We begin by supposing, as in Lemma 1, that at some v all blocking trades involve only proposals that are decided by minimal majority. Then, by minimality of the trades no more than one vote is ever traded on any given proposal (although trades could involve bundles of proposals). If i does not trade, then $\sigma_i(v') = \sigma_i(v)$, by construction. If i does trade, recall the notation used on the proof of Lemma 1 and call $\widetilde{P}$ the set of proposals on which i trades, $\widetilde{P}^{l(i)}$ the subset i wins pre-trade and loses post-trade, and $\widetilde{P}^{w(i)}$ the subset i loses per-trade and wins post-trade. By minimality, the resolution of all proposals on which votes are traded must change. Hence $\widetilde{P}^{l(i)} \cup \widetilde{P}^{w(i)} = \widetilde{P}$. Although the two sets may have different cardinality, by definition of pairwise improving trade, $\sum_{k\in\widetilde{P}^{l(i)}} x_i^k < \sum_{k\in\widetilde{P}^{w(i)}} x_i^k$. Since a single vote is traded on each proposal, we have:

$$
\begin{aligned}
\sigma_i(v') - \sigma_i(v) &= \left(\sum_{k\in\widetilde{P}^{w(i)}} x_i^k v_i'^k + \sum_{k\in\widetilde{P}^{l(i)}} x_i^k v_i'^k\right) - \left(\sum_{k\in\widetilde{P}^{w(i)}} x_i^k v_i^k + \sum_{k\in\widetilde{P}^{l(i)}} x_i^k v_i^k\right) \\
&\geq \sum_{k\in\widetilde{P}^{w(i)}} x_i^k - \sum_{k\in\widetilde{P}^{l(i)}} x_i^k \\
&> 0.
\end{aligned}
$$

The score of voter i has increased.

Hence if i trades, $\sigma_i(v') > \sigma_i(v)$. At any future step, 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 j whose score increases. The scores of all voters executing pairwise minimal trades on proposals decided by minimal majority must increase.

Suppose now that at v some blocking minimal trades involve proposals that are not decided by minimal majority. Call the set of such proposals $\widehat{P}(v)$. On such proposals, no single vote is pivotal, and hence trades must concern more than one vote. As a result, although $\sum_{k\in\widetilde{P}^{l(i)}} x_i^k < \sum_{k\in\widetilde{P}^{w(i)}} x_i^k$ must continue to hold by definition of payoff-improving trade, $\sigma_i(v') < \sigma_i(v)$ is possible (as in Example 2 in the text). But, by minimality, all proposals on which votes are traded must be decided by minimal majority after trade. Hence $|\widehat{P}(v')| < |\widehat{P}(v)|$, and since $|\widehat{P}(v)| \leq K < \infty$, blocking trades on non-minimal majority proposals can happen at most a finite number of times. Hence the logic of the proof of Theorem 1 applies here as well: for any R, the number of trades on non-minimal majority proposals must be finite, and because score functions are bounded and the number of voters is finite, so must be the number of trades on minimal majority proposals. Hence trading always ends after a finite number of steps, and a Pivot-stable allocation of votes always exists. Because the argument in the proof makes no restriction on R, the result holds for all R. ■

To prove Propositions 3 and 4, we exploit a result from the literature[14], restated in the following Lemma.

Lemma . *For any K, N, and z, the Condorcet winner, if it exists, can only be* $\mathbf{P}(v_0)$.

Proof. On any single proposal, the majority of the votes at v_0 reflect the preferences of the majority of the voters. For any number of proposals $m \in \{1, ..., K\}$, consider the outcome $\mathbf{P}(v_0, m^-)$ obtained by deciding m proposals in the direction favored by the minority at v_0, and the remainder $K-m$ in the direction favored by the majority. Consider the alternative outcome $\mathbf{P}(v_0, (m-1)^-)$, obtained by deciding one fewer proposal in favor of the minority at v_0. By construction, $\mathbf{P}(v_0, (m-1)^-)$ must be majority-preferred to $\mathbf{P}(v_0, m^-)$. Hence for any $m \in \{1, ..., K\}$, $\mathbf{P}(v_0, m^-)$ cannot be the Condorcet winner. But by varying m between 1 and K, $\mathbf{P}(v_0, m^-)$ spans all possible $\mathbf{P} \neq \mathbf{P}(v_0)$. Hence if the Condorcet winner exists, it can only be $\mathbf{P}(v_0)$. ■

Proposition 3. *If $N = 3$, then for all K and z, if the Condorcet winner exists, $\mathcal{P}(\mathcal{V}^*)$ is a singleton, and is the Condorcet winner.*

Proof. By Lemma 4, if the Condorcet winner exists, it can only be $\mathbf{P}(v_0)$. But then with $N = 3$ no trade can take place: if the Condorcet winner exists, v_0 cannot be blocked. Thus $\mathbf{P}(\mathcal{V}^*)$ equals $\mathbf{P}(v_0)$ and is the Condorcet winner. ■

Proposition 4. *If $K = 2$, then for all N and z, $\mathcal{P}(\mathcal{V}^*)$ is a singleton and is the Condorcet winner, if the Condorcet winner exists. If $\mathbf{P}(\mathcal{V}^*) \neq \mathbf{P}(v_0)$, a majority prefers $\mathbf{P}(\mathcal{V}^*)$ to $\mathbf{P}(v_0)$.*

Proof. Suppose, with no loss of generality, that $\mathbf{P}(v_0) = AB$–both proposals pass. All members of a blocking coalition must strictly gain from the trade. Hence any blocking trade must be such that both proposals change direction, because pivotal voters trading away their vote on one proposal must be compensated by moving to a winning position on the other proposal. It follows that along any path of trades the only two possible outcomes are AB, at $t = 0, 2, 4, ..,$ and $\varnothing$–both proposals fail–at $t = 1, 3, 5, ...$ We know from Theorem 1 that $\mathcal{P}(\mathcal{V}^*)$ is not empty. Hence $\mathcal{P}(\mathcal{V}^*) \subseteq \{AB, \varnothing\}$. Partition all voters into two sets of voters C_{AB} and $C_\varnothing$ where $i \in C_{AB} \iff AB \succ_i \varnothing$, that is, C_{AB} is composed of all voters who prefer AB to $\varnothing$; and $i \in C_\varnothing \iff \varnothing \succ_i AB$, that is, $C_\varnothing$ is composed of all voters who prefer $\varnothing$ to AB.[15] The two sets have cardinality N_{AB} and $N_\varnothing$, respectively. Note that blocking coalitions can only be formed within each set: for any path of trade, all members of a blocking coalition at t even must belong to $C_\varnothing$, and at t odd must belong to C_{AB}. Suppose first $N_{AB} > N_\varnothing$. Then at v_0, C_{AB} holds a total of N_{AB} on each proposal, and $C_\varnothing$ a total of $N_\varnothing$ votes, again on each proposal. Since $N_{AB} > N_\varnothing$, on each proposal voters in C_{AB} initially hold more votes than voters in $C_\varnothing$. Since blocking trades must always take place *within* either C_{AB} and $C_\varnothing$, this relation is true at every step of the trading path. But then $\varnothing$ cannot be a Pivot-stable outcome, because at any vote allocation v_t where $\mathbf{P}(v_t) = \varnothing$, $\varnothing$ is blocked by C_{AB}.

To see this, notice that because $\mathbf{P}(v_0) = AB$ and $N_{AB} > N_\varnothing$ it cannot be the case that all

[14] See Park (1967) and Kadane (1972).

[15] Ordering outcomes from most to least preferred, C_{AB} includes voters with rankings $\{\{AB, A, B, \varnothing\}, \{AB, B, A, \varnothing\}, \{A, AB, \varnothing, B\}, \{B, AB, \varnothing, A\}\}$; $C_\varnothing$ includes voters with rankings $\{\{\varnothing, A, B, AB\}, \{\varnothing, B, A, AB\}, \{A, \varnothing, AB, B\}, \{B, \varnothing, AB, A\}\}$.

voters in C_{AB} prefer $\varnothing$ to A or A would have failed at v_0. Similarly, it cannot be the case that all voters in C_{AB} prefer $\varnothing$ to B. Thus there must be at least one voter in C_{AB} who prefers B to $\varnothing$ and at least one voter in C_{AB} who prefers A to $\varnothing$. It follows that C_{AB} can always block v_t by giving all of its B votes to a voter in C_{AB} who prefers B to $\varnothing$ and giving all of its A votes to a voter in C_{AB} who prefers A to $\varnothing$. Because such a blocking trade is always possible, at any v_t reachable from v_0, it then follows that $\mathcal{P}(\mathcal{V}^*) = \{AB\}$. Identical logic shows that if $N_{AB} < N_{\varnothing}$, then $\mathcal{P}(\mathcal{V}^*) = \{\varnothing\}$. Because N is odd, $N_{AB} = N_{\varnothing}$ is impossible. Thus $\mathcal{P}(\mathcal{V}^*)$ must always be a singleton.

By Lemma 4, only AB can be the Condorcet winner. Because $\mathbf{P}(v_0) = AB$, it must be the case that AB is majority preferred to both A and B, and is the Condorcet winner if it is also majority preferred to $\varnothing$, i.e. if $N_{AB} > N_{\varnothing}$. But we just established that if $N_{AB} > N_{\varnothing}$, $\mathbf{P}(\mathcal{V}^*) = AB$. Hence if the Condorcet winner exists, $\mathbf{P}(\mathcal{V}^*)$ is the Condorcet winner. If $N_{AB} < N_{\varnothing}$, the Condorcet winner does not exist. In such a case, $\mathbf{P}(\mathcal{V}^*) = \varnothing$, and, since $N_{AB} < N_{\varnothing}$, $\mathbf{P}(\mathcal{V}^*)$ is majority preferred to $\mathbf{P}(v_0)$, concluding the proof of the proposition. ■

References

[1] Baron D. and J. Ferejohn, 1989, "Bargaining in Legislatures", *American Political Science Review*, 83, 1181-1206.

[2] Bernholz, P., 1973, "Logrolling, Arrow Paradox and Cyclical Majorities", *Public Choice*, 15, 87-96.

[3] Buchanan, J. and G. Tullock, 1962, *The Calculus of Consent*, Ann Arbor: University of Michigan Press.

[4] Casella, A., 2005, "Storable Votes", *Games and Economic Behavior*, 51, 391-419.

[5] Casella, A., A. Llorente-Saguer and T. Palfrey, 2012, "Competitive Equilibrium in Markets for Votes", *Journal of Political Economy*, 120, 593-658.

[6] Casella, A. and T. Palfrey, 2016, "An Experimental Study of Vote Trading", *Working Paper*, Columbia University.

[7] Casella, A. and T. Palfrey, 2017, "Trading Votes for Votes: A Dynamic Theory", *Working Paper*, California Institute of Technology.

[8] Coleman, J., 1966, "The Possibility of a Social Welfare Function", *American Economic Review*, 56, 1105-1122.

[9] Dal Bo, E., 2007, "Bribing Voters", *American Journal of Political Science*, 51, 789–803.

[10] Dekel, E., M. Jackson, and A.Wolinsky, 2008, “Vote Buying: General Elections”, *Journal of Political Economy*, 116, 351–380.

[11] Dekel, E., M. Jackson, and A.Wolinsky, 2009, “Vote Buying: Legislatures and Lobbying”, *Quarterly Journal of Political Science*, 4, 103–128.

[12] Downs, A., 1957, *An Economic Theory of Democracy*, New York: Harper.

[13] Downs, A., 1961. "Problems of Majority Voting: In Defense of Majority Voting", *Journal of Political Economy*, 69, 192-199.

[14] Feldman, A., 1973, "Bilateral Trading Processes, Pairwise Optimality, and Pareto Optimality", *Review of Economic Studies*, 40, 463-73.

[15] Feldman, A., 1974, "Recontracting Stability", *Econometrica*, 42, 35-44.

[16] Ferejohn, J., 1974, "Sour Notes on the Theory of Vote Trading", *Social Science Working Paper No.41*, California Institute of Technology.

[17] Gale D. and L. Shapley, 1962, "College Admissions and the Stability of Marriage", *American Mathematical Monthly*, 69, 9-15.

[18] Goeree J. and J. Zhang, 2017, "One man, one bid", *Games and Economic Behavior*, 101, 151-171.

[19] Green, J., 1974, "The Stability of Edgeworth's Recontracting Process", *Econometrica*, 42, 21-34.

[20] Groseclose T. and J. Snyder, 1996, “Buying Supermajorities”, *American Political Science Review*, 90, 303-315.

[21] Haefele, E., 1971. "A Utility Theory of Representative Government", *American Economic Review*, 61, 350-367.

[22] Hortala-Vallve R., 2012, "Qualitative Voting", *Journal of Theoretical Politics*, 24, 526-554.

[23] Jackson, M. and H. Sonnenschein, 2007, "Overcoming Incentive Constraints by Linking Decisions, *Econometrica*, 75, 241–257.

[24] Jackson, M. and A. Watts, 2002, "The Evolution of Social and Economic Networks", *Journal of Economic Theory*, 106, 265-295.

[25] Jackson, M. and A. Wolinsky, 1996, "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, 71, 44-74.

[26] Kadane, J., 1972, "On the Division of the Question", *Public Choice,* 13, 47-54.

[27] Koehler D., 1975, "Vote Trading and the Voting Paradox: A Proof of Logical Equivalence", *American Political Science Review,* 69, 954-960.

[28] Lalley, S. and G. Weyl, 2016, "Quadratic Voting", https://ssrn.com/abstract=2003531.

[29] Miller, N., 1977, "Logrolling, Vote Trading and the Paradox of Voting: A Game Theoretical Overview", *Public Choice*, 30, 51-75.

[30] Moulin, H., 1988, *Axioms of Cooperative Decision Making*, Cambridge, UK: Cambridge University Press.

[31] Myerson, R., 1993, "Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems", *American Political Science Review*, 87, 856-869.

[32] Park, R.E., 1967, "The Possibility of a Social Welfare Function. Comment", *American Economic Review*, 57, 1300-1304.

[33] Philipson, T. and J. Snyder, 1996, "Equilibrium and Efficiency in an Organized Vote Market", *Public Choice*, 89, 245-265.

[34] Riker, W. and S. Brams, 1973, "The Paradox of Vote Trading", *American Political Science Review*, 67, 1235-1247.

[35] Roth, A. and M. Sotomayor, 1990, *Two-Sided Matching: A Study in Game-Theoretic Modelling and Analysis*, Cambridge, UK: Cambridge University Press.

[36] Roth A. and J. Vande Vate, 1990, "Random Paths to Stability in Two-Sided Matching", *Econometrica*, 58, 1475-1480.

[37] Schwartz, T., 1975, "Vote Trading and Pareto Efficiency", *Public Choice*, 24, 101-110.

[38] Stratmann, T., 1992, "The Effects of Logrolling on Congressional Voting", *American Economic Review*, 82, 1162-1176.

[39] Tullock, Gordon, 1959, "Problems of Majority Voting", *Journal of Political Economy*, 67, 57l-79.

[40] Tullock, Gordon, 1961, "Problems of Majority Voting: Reply to a Traditionalist", *Journal of Political Economy*, 69, 200-203.

[41] Tullock, Gordon, 1970, "A Simple Algebraic Logrolling Model", *American Economic Review*, 60, 419-426.

[42] Watts, A., 2001, "A Dynamic Model of Network Formation", *Games and Economic Behavior*, 34, 331-341.

[43] Wilson, R., 1969, "An Axiomatic Model of Logrolling", *American Economic Review,* 59, 331-341.

[44] Xefteris D. and N. Ziros, 2017, "Strategic Vote Trading in Power Sharing Systems", *American Economic Journal: Microeconomics*, 9, 76-94.