Trading Votes for Votes: A Dynamic Theory[1]

Alessandra Casella[2] Thomas Palfrey[3]

March 1, 2017

[1]We thank Enrico Zanardo, Kirill Pogorelskiy and Manuel Puente for research assistance, and participants at several research conferences and seminars for their comments. We especially thank Micael Castanheira, Michel LeBreton, Cesar Martinelli, Debraj Ray, Richard Van Weelden, Rajiv Vohra, and Alistair Wilson for detailed comments and suggestions. 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 propose a framework to study sequential rounds of vote trading in which voters exchange ballots over different separable binary proposals. We ask: (1) Does decentralized vote trading converge to a stable allocation of votes? (2) What properties must this allocation possess? We prove that if trade is constrained to be pairwise, a stable allocation is always reached in a finite number of trades, for any number of voters and issues, and for any separable preferences. The result however does not extend to trading coalitions of arbitrary size. Conversely, if coalitions have arbitrary size, the stable outcome, if it exists, must be Pareto optimal, but this need not be true if trade is restricted to be pairwise. For any size of the trading coalition, there are special cases such that the stable outcome exists and must coincide with the Condorcet winner, if the latter exists. In general however, existence and welfare properties of the stable outcome have no logical link to the existence of the Condorcet winner. If trading is farsighted, the properties of vote trading are not any stronger.

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. If we think of the trades as a sequential dynamic process, 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 open up new trading possibilities. What is the most productive approach to modeling such a process?

This paper is inspired by the similarity between the logic of sequential rounds of vote trading and other games studied in the literature in which convergence to a stable allocation occurs, if it does, in the absence of a price adjustment. 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 and deleting links in the formation of a network (Jackson and Wolinsky, 1996; Watts, 2001; Jackson and Watts, 2002), or in sequences of barter trades in an exchange economy without money (Feldman, 1973 and 1974; Green, 1974). 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 simple rule–in essence an algorithm–specifying the steps on the dynamic path. An alloca-

[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.

tion is stable if there exists no coalition that can, on its own, achieve an alternative allocation preferred by all of its members; a move along the path is made when such a blocking coalition exists and implements a preferred allocation. Using such a framework, we ask whether a stable allocation of votes exists, whether sequential trading converges to a stable allocation, and what are the properties of the ultimate stable vote allocations that arise.

The desirability of such an approach was already clear to the early literature on vote trading. Our definition of stability echoes the definition proposed by Park (1967), and our trading algorithm, if specialized to pairwise trading only, is related to the sequence of trades posited by Riker and Brams (1973) and Ferejohn (1974). However, contributors to the early literature 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 potential of the approach was not realized.

The group decision-making problem we study is defined by an odd number of voters facing several binary proposals, each of which may pass or fail; after vote trading agreements are made, each proposal is decided by majority rule.[2] Every committee member can be in favor or opposed to any proposal and has separable preferences across proposals.

A vote trade is a commitment by some voters to cast their votes on certain issues against their preferences, in exchange for a similar commitment by other voters on some other issues. Formally it is modeled as an enforceable contract among a subset of voters of arbitrary size.

Dynamic trading processes are formulated as a family of algorithms. Given an allocation of votes, a selection rule, possibly random, selects one of the existing blocking coalitions and a payoff-improving exchange of votes. This defines a new allocation of votes, and the algorithm again selects a new coalition and improving trade from the set of all improving trades available at this new vote allocation. The algorithm continues until a vote allocation is reached where there are no more improving trades for any coalition. The family of these algorithms is populated by considering all possible selection rules, whenever multiple payoff-improving trades are possible. As remarked in Riker and Brams (1973), the requirement that a vote

[2] The approach can be extended to general voting rules.

trade be strictly welfare improving for all traders implies that the votes being traded must be pivotal, and thus we call our family of algorithms the *Pivot algorithms.*

Our first result, contradicting a conjecture by Riker and Brams, is that if trade is restricted to be pairwise, then a Pivot algorithm *always* generates a stable vote allocation in a finite number of steps, for any number of voters, any number of proposals, any configuration of (separable) preferences, and any selection rule. This result is very general and mostly unexpected: stability is achieved without any of the restrictions that have proved necessary for convergence in decentralized matching, or in network formation, or in barter trade. Even with vote trading, however, convergence is not guaranteed if coalitions can be of arbitrary size: we construct an example where the vote allocation can cycle and trading need never end.

If a stable vote allocation is reached, what welfare properties will it possess? The early literature on vote trading 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.[3] The conjecture was rejected by Riker and Brams' (1973) influential "paradox of vote trading": if vote trading is pairwise and binding, there are non-pathological preference profiles such that each pair of voters individually gain from trading their votes and yet everyone strictly prefers the no-trade outcome to the final outcome of trading. Opposite examples where vote trading is Pareto superior to no-trade can easily be constructed as well,[4] and the literature eventually ran dry with the tentative conclusion that no general statement on the desirability of vote trading can be made.

Our analysis confirms Riker and Brams' result– it is indeed possible for pairwise trading to lead to Pareto inferior outcomes. However, in contrast to this negative finding, we also show that if trading coalitions can take any arbitrary size, then if a stable allocation is reached it *must* be Pareto optimal, and again this must hold for any number of voters or issues, for any realization of preferences, and for any rule selecting among possible trades.

In special cases we can say more. In particular, when the committee is faced with

[3]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 (1970), Tullock (1970), and Wilson (1969).

[4]For example, Schwartz (1975).

only two proposals (and thus, since each proposal can either pass or fail, four possible outcomes), then for any number of voters and any preference profile, 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 regardless of whether trade is restricted to be pairwise or if arbitrary coalitions are allowed. They are somewhat surprising, as it has always been understood that vote trades' ambiguous welfare properties are due to the externalities inherent in the exchanges. But externalities are clearly present in the two-proposal case, and still the Pivot algorithms always deliver outcomes with desirable welfare properties.

Regardless of coalition size, convergence to the Condorcet winner, if it exists, is also guaranteed if the group size limited to three, for any number of proposals. But the result does not extend to more than two proposals and more than three voters: regardless of coalition size, the Pivot algorithms need not converge to the Condorcet winner, if there are four or more voters and three or more proposals.

Taken together, our results also address a second central debate in the early literature: the relationship between stability and the existence of the Condorcet winner. Buchanan and Tullock (1962) and Coleman (1966) conjectured that vote trading may offer the solution to majority cycles in the absence of a Condorcet winner. Starting with Park (1967), a number of authors studied and rejected the conjecture[5], but again the different scenarios and the incompletely specified trading rules make comparisons difficult. In our model, under the assumptions of issue-by-issue voting and binding trades, there is no logical connection between stability of a vote allocation and existence of the Condorcet winner. We prove that a coalition-stable vote allocation may exist in the absence of a Condorcet winner, may not exist when the Condorcet winner exists, or it may exist and yet differ form the Condorcet winner.

The stability notion discussed so far, as well as the Pivot algorithms, implicitly assume that voters do not take into account future trades when evaluating the benefits of a current trade. The final section of the paper explores the theoretical implications of farsighted vote trading. We do so by extending the model to allow

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

voters to take into account the entire future path of trades. While we use a different definition of farsighted stability, our analysis is similar in spirit to recent approaches in cooperative game theory that explore the implications of forward looking sophistication (Chwe (1994), Dutta and Vohra (2015), Ray and Vohra (2015)).[6]

As in this literature, we define farsighted dominance in terms of a sequence of vote trades, beginning at one vote allocation v and ending at some other vote allocation v'. At each step of the sequence, all members of the coalition involved in the next trade strictly prefer the outcome under v' to the outcome that would correspond to the current vote allocation. Hence, all myopically improving trades are one-step farsighted trades, but the trades in a farsighted sequence are not necessarily myopically improving. The farsighted core in the vote trading game consists of the set of vote allocations that are not farsightedly dominated by any other vote allocation. Thus the farsighted core is the natural farsighted parallel to the myopic notion of Pivot stability, and farsighted domination sequences are the farsighted parallel to the myopic trading algorithms. We are interested in the existence and properties of vote allocations in the farsighted core reachable via domination chains *from the initial vote allocation.*[7]

If such farsightedly stable vote allocations exist, they must be Pareto optimal. On the whole, however, farsightedness does not lead to better properties for vote trading. We show that while the farsighted core is always non-empty, reaching it from the initial vote allocation may be impossible. What is more surprising, we find that under farsightedness achieving the Condorcet winner is possible only if vote trading does *not* take place, and simple examples exist where the Condorcet winner exists and is always reached under myopic trading, but the farsighted stable vote allocation reached by farsighted trading delivers a different outcome. Again this counters a conjecture from the early literature, where forward-looking behavior was not modeled explicitly but was believed to select the Condorcet winner whenever it

[6]This is distinct from a non-cooperative game approach, where the dynamic process is modeled as a predetermined extensive form game. Such an approach would require a much more rigid specification of the details and timing of play.

[7]The literature has proposed alternative definitions of farsighted stability, with a focus on solution concepts that extend the von Neumann - Morgenstern solution to allow for farsighted domination (in addition to the authors cited above, see for example Diamantoudi and Xue (2003) and Mauleon et al. (2011)). We discuss the relationship between our solution and alternative approaches in Appendix B.

exists.[8]

After describing the model and specifying definitions and results under myopia (Section 2), and under farsightedness (Section 3), Section 4 summarizes our conclusions and discusses possible directions of future research.

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 $P = \{1, ..., k, ..., K\}$. Committee members have separable preferences represented by a profile of 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. We normalize to 0 the value of any proposal failing.[9] 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 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 $v = (v_1, ..., v_N)$ a *vote allocation*, i.e., a profile of vote holdings for all voters and proposals. The initial vote allocation is denoted by v_0, and we set $v_0 = (\mathbf{1}, \mathbf{1}, ..., \mathbf{1})$: each voter is initially endowed with one vote on each proposal. Let $\mathcal{V}$ denote the set of feasible vote allocations: $v \in \mathcal{V} \Longleftrightarrow \sum_i v_i^k = N$ for all k and $v_i^k \geq 0$ for all i, k.[10] A *trade* is any pair of vote allocations (v, v'), such that $v, v' \in \mathcal{V}$ and $v \neq v'$.

Given a feasible vote allocation v, when voting takes place, each voter has a dominant strategy to cast all his votes in favor of the proposal if his proposal's

[8] See for example Park (1967).

[9] Although it is convenient to work with the profile of cardinal values z, our analysis exploits ordinal rankings only.

[10] Note that $\sum_k v_i^k \neq K$ is feasible because we do not restrict trades to be one-to-one. Of course, the aggregate constraint $\sum_i \sum_k v_i^k = NK$ must hold.

value is positive ($z_i^k > 0$), and against the proposal if his 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 $(N+1)/2$ 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: $u_i(v) = \sum_{k \in \mathbf{P}(v)} z_i^k$. Preferences over outcomes are assumed to be strict. That is, $u_i(v) = u_i(v')$ if and only if $\mathbf{P}(v) = \mathbf{P}(v')$.[11]

Our focus is on the existence and properties of vote allocations that hold no incentives for trading. Consider any trade from v to v', and let $d_i'^k(v, v') = |v_i'^k - v_i^k|$ denote the absolute change in vote holdings for individual i on proposal k. We define:

Definition 1 *(v, v') is a **strictly payoff improving trade** if $v' \in \mathcal{V}$ and, for all i, $d_i(v, v') \neq 0 \Rightarrow u_i(v') > u_i(v)$.*

That is, a trade is strictly payoff improving if every voter who is involved in the trade is strictly better off with the outcome generated by the new vote allocation, and the new vote allocation is feasible. We then say:

Definition 2 *A coalition of voters $S = \{i, i', i'', ..\}$, is said to **block** v if at v there exists a strictly payoff improving trade (v, v') for all $i \in S$.*

Definition 3 *A vote allocation $v \in \mathcal{V}$ is **stable** if there exists no coalition of voters who block v.*

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:

[11] Some of the examples later in the paper allow voters to have weak preferences. This is done for expositional clarity, and the examples are easily modified to strict preferences.

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

We allow S to have any arbitrary size between 2 and N. The older literature, however, made most progress when restricting trade to be pairwise and argued with some plausibility that the difficulty of organizing a coalition makes pairwise trading the empirically relevant case. We will discuss explicitly when restricting trade affects the theoretical results.

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 begin with some definitions.

For any given allocation of votes, v, denote by $\mathcal{V}^+(v)$ the set of trades that can be executed by some blocking coalition (we use the superscript $+$ to indicate that the trades must be strictly payoff improving).

Definition 4 $(v, v'') \in \mathcal{V}^+(v)$ *is a* ***reduction*** *of* $(v, v') \in \mathcal{V}^+(v)$ *if for all* i, k, $d_i^k(v, v') = 0 \Rightarrow d_i^k(v, v'') = 0$, *and for all* i, k, $d_i^k(v, v') \geq 0 \Rightarrow d_i^k(v, v') \geq d_i^k(v, v'')$, *with* $d_i^k(v, v') > d_i^k(v, v'')$ *for some* i, k.

Definition 5 *Consider trade* (v, v') *executed by coalition* S. *We say that trade* (v, v') *is a* ***minimal*** *trade if the following two conditions hold: (1) there does not exist a reduction of* (v, v')*; (2) there does not exist a proper subcoalition* $C \subset S$ *that blocks* v.

By concentrating on minimal trades, we require that trade sequences consist of elementary trades: at each step, traders will exchange the minimal number of votes required to achieve a payoff improvement, for a given coalition, and the trading coalition will be no larger than necessary. If a trade is minimal, then it does not include redundant votes (votes that have no effect on the outcome); it does not bundle together multiple payoff improving trades, and it does not include traders whose presence is not required for strict gains by the remainder of the coalition. Suppose for example that voters 1 and 2 profit from exchanging votes on proposals A and B. Then: (1) we do not allow one or both of them to add to the trade an extra non-pivotal vote; (2) if 1 and 2 also profit from exchanging votes on proposals C and D, we require that they execute the two trades in two steps; (3) we do not allow voter 3 to participate in the trade, for instance as an intermediary, even if 3 too benefits from changing the direction of A and B; (4) we privilege the trade between 1 and 2 on proposals A and B over an alternative trade on other proposals that does require 3's presence.

Minimality is important because the dynamic path of the sequential trades depends on the vote allocation at each step. Including redundant votes or voters and bundling trades can affect the dynamic path, even when it is myopically irrelevant.

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

Definition 6 *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 minimal strictly improving trade, stop. If there is one such trade, execute it. If there are multiple such trades, choose one according to a choice rule* R*. Continue in this fashion until no further minimal strictly improving trade exists.*

The definition describes a family of algorithms, depending on the choice rule R that is applied when multiple minimal trades are possible. Rule R specifies how the algorithm selects among them; for example, R may select each possible trade with equal probability; or give priority to trades with higher total gains; or to trades involving specific voters. Note that rule 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 satisfy minimality. The family of Pivot algorithms corresponds to the class of possible R rules, and individual algorithms differ in the specification of rule R. [12]

Payoff improving trades are not restricted to two proposals only, nor to exchanging one vote for one vote: a voter can trade his 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 restrictions we are imposing are that trades be minimal and strictly payoff-improving for all traders. If a trade is payoff improving and minimal, it is a legitimate trade under the Pivot algorithms.

[12] At this point, it is not necessary to be more specific about R.

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

Lemma 1 *(Riker and Brams) In any Pivot algorithm, all votes transferred must be pivotal.*

Proof. Immediate by definition of Pivot algorithm. All trades selected by any Pivot algorithm must be minimal. But then all votes traded must affect the outcome, and thus be pivotal. Suppose not. Then there exists a traded vote whose exchange does not modify the outcome. But then there exists a reduction of the trade in which the vote is not exchanged.□

2.2 Pivot-stable vote allocations

Do Pivot algorithms converge to stable vote allocations?[13] The question is not trivial because any Pivot trade changes outcomes and alters the existing set of payoff-improving trades, potentially leading to new Pivot trades, in a sequence that in theory may well result in a perennial cycle. We define:

Definition 7 *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, following rule R.*

We find:

Theorem 1 *A Pivot-stable allocation of votes exists for all K, N, z, and R if trade is restricted to be pairwise. If coalitions can have arbitrary size, nonexistence of Pivot-stable vote allocations is possible.*

Proof. The proof is in two steps. We show first that restricting trade to be pairwise is sufficient to ensure Pivot-stability, for all K, N, z, and R. We then show that, without this restriction, Pivot-stable vote allocations can fail to exist.

(1) Suppose only pairwise trades are allowed. Then, by Lemma 1 and $v_0 = \{\mathbf{1}, \mathbf{1}, ..\}$, if a trade occurs at v_0 it can only concern proposals that at v_0 are decided

[13]Stated differently, do sequential myopic trades converge to the core?

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 $t > 0$. Since $v_0 = \{\mathbf{1}, \mathbf{1},..\}$, it follows that 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:

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

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 $\{k, k', ..\}$ and increase by one vote on some other proposals $\{\widetilde{k}, \widetilde{k}', ..\}$. 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 pairwise improving trade, $\sum_{k \in \mathbf{P}_{i,t}^-} x_i^k < \sum_{k \in \mathbf{P}_{i,t}^+} x_i^k$ and, since a single vote is traded on each proposal, $\sum_{k \in \mathbf{P}_{i,t}^-} x_i^k v_{it}^k < \sum_{k \in \mathbf{P}_{i,t}^+} x_i^k v_{it+1}^k$. 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.[14] 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.

(2) We now prove that coalitional trading can lead to nonexistence of Pivot-stable allocations for some K, N, z, and R. 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.

At v_0, all proposals pass, 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

[14]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 $\left\lceil \frac{K(K-1)}{2} \right\rceil \left\lceil \frac{(N-1)}{2} \right\rceil$.

	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 1: *With arbitrary coalition size, Pivot may never converge. An example.*

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 the coalition is minimal (no subcoalition has a strict payoff-improving trade), and v_1 cannot be reached by the coalition by trading fewer votes. But note that 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 $\mathbf{P}(v_2) = \{A, B\}$, 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: $\mathbf{P}(v_3) = \{A, B, C, D\}$, 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. □

With pairwise trading, the generality of the result in the theorem is surprising. Guaranteeing convergence without any restriction on the selection rule R has no counterpart either in matching, (Roth and Vande Vate, 1990; Diamantoudi et al., 2004), or networks (Watts, 2001; Jackson and Watts, 2002), or barter trade (Feldman, 1973), all models based on pairwise interactions and yet requiring some randomness in R to ensure that any cycle will be broken. In vote trading, Riker and Brams (1973) conjectured that convergence required limiting the number of allowed trades per proposal, but the theorem shows that their conjecture does not hold. 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.[15]

[15] Of separate interest is the observation that with pairwise trading Pivot stability is guaranteed for any arbitrary initial allocation of votes v_0, not only $v_0 = \{\mathbf{1}, \mathbf{1}, ..\}$. The result does not

It is then surprising to find that the result does not extend to coalitions of arbitrary size. In the presence of coalitions, stability may fail because trades can be profitable for the coalition even when the pairwise trades that are part of the overall exchange are not: coalition members benefit from the positive externalities that originate from the trades of other members. As a result, the score function is no longer monotonically increasing in the number of trades. In the example used in the proof, voters $1, 2, 3$, and 4 have a score of 5 before the coalition trade and a score of 4 after the trade. Cycles become possible, and stability need not be achieved.

2.3 Preferences over Pivot-stable outcomes

Definition 8 *An outcome* $\mathbf{P}(v)$ *is a* ***Pivot-stable outcome*** *if* v *is a Pivot-stable vote allocation.*

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

What are the welfare properties of $\mathcal{P}(\mathcal{V}_R^*)$? Our institution-free approach demands a welfare evaluation that is equally institution-free. We ask whether outcomes in $\mathcal{P}(\mathcal{V}_R^*)$ 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. We begin with the following result:

Theorem 2 *If coalitions can have arbitrary size, then if a Pivot-stable outcome exists it must be in the Pareto set, for all* K*,* N*,* z*, and* R*. If trade is restricted to be pairwise, there exist* K*,* N*,* z*, and* R *such that no Pivot-stable outcome is in the Pareto set.*

---

follow immediately because with arbitrary v_0 it is possible to have trades on proposals that are not decided by minimal majority. And if a proposal is not decided by minimal majority, an initial minimal trade can involve more than a single vote. As a result, a trader's score function may decrease even if the trade is strictly payoff-improving. However, by minimality, the first trade involving a given proposal *must* bring the proposal to minimal majority. All further trades on that proposal must then consist of a single vote. Hence with K proposals, for any v_0 there are at most $K-1$ trades for which the score may fall. For all other trades the score must increase. This is enough to establish that trade must end in finite time.

Proof. We first show that allowing trade among coalitions of arbitrary size is a sufficient condition for Pareto optimality, if a Pivot-stable outcome exists. We then show that Pareto-optimality can break down with pairwise trading.

(1) Suppose that arbitrary coalitions are allowed and 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.[16]

(2) The example in Table 2 shows that Pareto optimality need not hold when trade is pairwise. Consider the following matrix, with $K = 5$ and $N = 5$. 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	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 2: *When trade is restricted to be pairwise, no outcome in the Pareto set may be Pivot stable. An example.*

When trade is restricted to be pairwise, no voter is pivotal, and thus v_0 cannot be blocked. The unique stable outcome is $\mathbf{P}(\mathcal{V}_R^*) = \mathbf{P}(v_0) = \{\varnothing\}$: all proposals fail. Yet, all proposals failing is not Pareto optimal: it is Pareto-dominated by all proposals passing. □

Can we say anything more precise? Our first set of answers is unexpectedly positive. To establish them, we exploit one result from the literature[17]. We report it here.

Lemma 2 *(Park and Kadane). If the Condorcet winner exists, it can only be* $\mathbf{P}(v_0)$.

[16] If the coalition trade is not minimal, it can be made so by eliminating redundant trades or traders.

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

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

This immediately implies Proposition 2:

Proposition 2 *If $N = 3$, then for all K, z, and R, if a Condorcet winner exists, $\mathcal{V}_R^*$ is always nonempty, $\mathcal{P}(\mathcal{V}_R^*)$ is a singleton, and $\mathbf{P}(\mathcal{V}_R^*)$ is the Condorcet winner.*

Proof. By Lemma 2, if the Condorcet winner exists, it can only be $\mathbf{P}(v_0)$. But then no trade can take place: if $N = 3$ and the Condorcet winner exists, v_0 cannot be blocked. Thus $\mathbf{P}(\mathcal{V}_R^*)$ equals $\mathbf{P}(v_0)$ and is the Condorcet winner.□

There is another scenario in which the Pivot-stable outcome is related to majority preferences:

Proposition 3 *If $K = 2$, then, for all N, z, and R: (1) $\mathcal{V}_R^*$ is always nonempty and $\mathcal{P}(\mathcal{V}_R^*)$ is a singleton.*[18] *(2) $\mathbf{P}(\mathcal{V}_R^*)$ is Pareto optimal. (3) If a Condorcet winner exists, then $\mathbf{P}(\mathcal{V}_R^*)$ is the Condorcet winner. (4) If $\mathbf{P}(\mathcal{V}_R^*) \neq \mathbf{P}(v_0)$, then a majority prefers $\mathbf{P}(\mathcal{V}_R^*)$ to $\mathbf{P}(v_0)$.*

Proof. See Appendix A.

Proposition 3 holds whether trade is restricted to be pairwise or allowed to involve coalitions of any size. It is interesting because it highlights the fact that Pareto suboptimality 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 algorithms is always Pareto optimal. The link between Pareto suboptimality and voting externalities is a central lesson from Riker and Brams (1973), but Pivot algorithms implement the same trading rule responsible for the vote-trading paradox in

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

that work, and yet with $K = 2$ vote trading performs well, both in a Pareto sense and in terms of majority preferences.[19]

However, the results from Pivot trading are more ambiguous if $N > 3$ and $K > 2$. Specifically:

Proposition 4 *If $K > 2$ and $N > 3$, then there exist z such that the Condorcet winner exists, but the Condorcet winner is not an element of $\mathcal{P}(\mathcal{V}_R^*)$ for any R.*

Proof. Consider the following example, with $K = 3$ 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 3: *Preference profile such that the Condorcet winner is not Pivot-stable.*

For the preference profile in Table 3, $\mathbf{P}(v_0) = \{ABC\}$ is the Condorcet winner. However, regardless of R, there is a unique Pivot-stable outcome, $\{A\}$. All minimal blocking coalitions in this example turn out to be pairwise, but pairwise trading is not an imposed restriction. There are three different possible trade chains, all stopping 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)[20], and, in parenthesis, the outcome corresponding to that allocation of votes, we can describe the three chains as $\{\{13cb(A),\ 45bc(ABC),\ 23ab(C),\ 45ca(A)\},\ \{23ab(C),\ 45ba(ABC),\ 13cb(A)\},$ and $\{23ab(C),\ 45ca(A),\ 45bc(ABC),\ 13cb(A)\}$.□

Proposition 4 immediately implies:

Corollary 1 *If $K > 2$ and $N > 3$, then there exist z such that for all R there exists some Pivot-stable vote allocation $v^* \in \mathcal{V}_R^*$ such that $\mathbf{P}(v^*) \neq \mathbf{P}(v_0)$, and $\mathbf{P}(v_0)$ is majority preferred to $\mathbf{P}(v^*)$.*

[19] 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.

[20] For example, $13cb$ indicates that voter 1 acquires a C vote from voter 3, in exchange for a B vote.

	1	2	3	4	5	6	7	8	9
A	2	-1	-1	-1	1	1	1	-1	3
B	-1	2	-1	-1	1	1	1	3	-1
C	-1	-1	2	-1	1	1	1	-1	3
D	-1	-1	-1	2	1	1	1	3	-1

Table 4: *Existence of the Condorcet winner does not guarantee convergence to a stable outcome. An example.*

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 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.

Our results thus say that the Pivot-stable outcome, when it exists, may or may not coincide with the Condorcet winner, when this too exists: it must in some cases ($K = 2$, $N = 3$), and need not in others. We can also take a step backward, and ask whether the existence of a Condorcet winner has any implication for the existence of a stable outcome, one of the central early debates on vote trading. As we show in Table 4, it is not difficult to modify the matrix in Table 1 so that $\{A, B, C, D\}$ is now the Condorcet winner.

Yet the argument used in the proof of Theorem 1 continues to apply identically: with arbitrary coalitions size, the existence of a Condorcet winner does not guarantee stability. In our model, with binding vote trades and proposal-by-proposal voting, there is no logical connection between the existence and welfare properties of a stable outcome and the existence of a Condorcet winner.

Social choice theorists have developed other useful characterizations of sets of outcomes that can emerge from voting. Prominent among them is the *Uncovered Set*, the set of all outcomes that are *not covered* by any other. Using the notation $\mathbf{P} \succ \mathbf{P}'$ to indicate that outcome $\mathbf{P}$ is preferred to $\mathbf{P}'$ by a majority of voters, the covering relation is defined as follows:

Definition 9 $\mathbf{P}$ covers $\mathbf{P}'$ *if and only if* $\mathbf{P} \succ \mathbf{P}'$ *and* $\mathbf{P}' \succ \mathbf{P}'' \Longrightarrow \mathbf{P} \succ \mathbf{P}''$.

When the Condorcet winner exists, the Uncovered set is a singleton and corresponds to the Condorcet winner.

Is there any relation between the Uncovered set and $\mathcal{P}(\mathcal{V}_R^*)$, the set of Pivot-stable outcomes? In general, the answer is negative: $\mathcal{P}(\mathcal{V}_R^*)$ can correspond to the Uncovered set, but can also be a strict subset, or a strict superset, or be fully disjoint. We know from Proposition 3 that with $K = 2$, $\mathcal{P}(\mathcal{V}_R^*)$ is always a singleton and corresponds to the Condorcet winner if the Condorcet winner exists. Thus if $K = 2$ and the Condorcet winner exists, $\mathcal{P}(\mathcal{V}_R^*)$ corresponds to the Uncovered set. However, if the Condorcet winner does not exist, the Uncovered set cannot be a singleton, and $\mathcal{P}(\mathcal{V}_R^*)$ must be a strict subset of the Uncovered set.[21] With $K > 2$ and $N > 3$, $\mathcal{P}(\mathcal{V}_R^*)$ may be empty. But even if trade is restricted to be pairwise (and thus $\mathcal{P}(\mathcal{V}_R^*)$ is not empty), it is not difficult to construct examples where the Condorcet winner exists, but multiple outcomes are Pivot -stable, and $\mathcal{P}(\mathcal{V}_R^*)$ is a superset of the Uncovered set.[22] In addition we know from Proposition 4 that with $K > 2$ and $N > 3$, the Condorcet winner need not belong to $\mathcal{P}(\mathcal{V}_R^*)$: as in the example in Table 3, the two sets can be fully disjoint.

3 Farsighted Vote Trading

As in most theoretical work on network formation, barter, and matching, the dynamic process we have studied so far is defined by a myopic algorithm: the Pivot algorithm is explicitly myopic. A natural question is whether the model can be extended to accommodate forward looking behavior by the voters. As we showed in a number of examples, strictly improving myopic trade can trigger subsequent trades by others that harm the initial traders, not only undoing their original gain but leading to a worse outcome than the pre-trade vote allocation.

One approach to modeling forward looking sophistication would be to reformulate the model as a dynamic extensive form game, and characterize the properties of the perfect equilibria of the game. This would require a different framework, one that would impose much more structure on the basic vote trading process–specifying

[21] Suppose $\mathbf{P}(v_0) = \{A, B\}$. $\{A, B\}$ is always majority preferred to both $\{A\}$ and $\{B\}$, but if the Condorcet winner does not exist, $\varnothing$ must be majority preferred to $\{A, B\}$. Hence $\{A, B\}$ and $\varnothing$ must belong in the Uncovered set. $\mathbf{P}(\mathcal{V}_R^*)$ on the other hand is a singleton and must correspond to either $\{A, B\}$ or $\varnothing$.

[22] We report one such example in Appendix A.

a well-defined sequence of moves, information sets, rationing rules. A more tractable approach, based on cooperative game theory, delivers a natural extension of the myopic model. model comes from cooperative games. The problem remains complex, perhaps even moreso. Because of the externalities involved and because the opportunities for trade depend on the vote allocation, vote trading cannot be represented under any of the existing cooperative models of farsightedness.[23] However, we show in this section that the Pivot algorithms lend themselves to a natural extension, which allows us to establish some initial results.

3.1 Farsighted stability

We begin with some preliminary conventional definitions. Given two vote allocations v and v', a coalition S is *effective* for (v, v') if $v' \in \mathcal{V}$ (v' is feasible) and $v'_i = v_i$ for all $i \notin S$. That is, voters in S can move the vote allocation from v to v' by reallocating votes among themselves only. A *chain* from v to v' is a collection of vote allocations $v^0, v^1, ..v^m$, with $v^0 = v$ and $v^m = v'$, and a corresponding collection of effective coalitions $S^1, .., S^m$ such that for all $t = 0, ..m-1$, S^{t+1} is effective for (v^t, v^{t+1}). A chain is a *farsighted chain* (an *F-chain*) if, in addition, $u_j(v^t) < u_j(v')$ for all $t = 0, ..m-1$, and all $j \in S^{t+1}$, i.e. if all members of all effective coalitions in the chain strictly prefer the final vote allocation to the allocation at which they trade. Coalitions in an F-chain thus differ from our earlier definition of blocking coalitions under two dimensions: (1) at any t, the members of coalition S^t, effective for (v^{t-1}, v^t) need not prefer v^t to v^{t-1}, either strictly or weakly; (2) they must however strictly prefer the final allocation v' to v^{t-1}.

In principle, we could constrain all effective coalitions on a chain to be formed of two voters only, and talk of pairwise trade. But the constraint and the term would be misleading: the logical basis of a farsighted trade, whether pairwise or not, is the farsighted understanding of the chain–it is the final allocation that matters–and hence unless the chain can be condensed into a single pairwise exchange, the relevant sum of all trades necessarily involves multiple voters. Requiring any individual trade to involve two voters only, but in the knowledge that other voters will also later trade, does not capture the concern about difficulties of coordination that had drawn the

[23]See for example Chwe (1994), Mauleon et al. (2011), Ray and Vohra (2015), Dutta and Vohra (2015), and the references therein.

older literature to concentrate on pairwise trades. The distinction between pairwise trading and trading among coalitions of arbitrary sizes loses its significance with farsightedness, and we drop it in this section: we always allow coalitions to be of arbitrary size at all steps of a farsighted chain.

The restriction to minimal trades and minimal coalitions is also less compelling in the farsighted case. Voters in a coalition may wish to exchange extra votes in order to prevent future adverse vote trades by others. This could be accomplished, for example, by forming a coalition larger than the minimal effective coalition, or by trading non-pivotal votes and cumulating votes on one issue in the hands of one member of the coalition.[24] In the purely myopic case, strategies such as these have no role, but excluding them would be artificially limiting in the farsighted case. Therefore, we dispense with the minimality requirements for trades when voters are allowed to be forward looking.

For any pair of vote allocations, v and v', v' is said to *farsightedly dominate (F-dominate)* v if there exists an F-chain from v to v'. Let $D(v) \equiv \{v' \in \mathcal{V} | v'$ F-dominates $v\}$. That is, $D(v)$ is the set of feasible vote allocations reachable from v via a chain of farsighted trades.

As noted earlier, the definition of stability we have used so far corresponds to the (myopic) core. The most immediate extension of our approach is to define farsighted stability by reference to the farsighted core:

Definition 10 *The* farsighted core*, C_F, is the set of all F-undominated vote allocations. That is, $C_F = \{v | \ D(v) = \varnothing\}$.*

Definition 11 *A vote allocation $v \in \mathcal{V}$ is* farsightedly stable *(F-stable) if and only if $v \in C_F$*

A vote allocation that is not myopically stable (in the sense of Definition 3) is not farsightedly stable. Hence the set of farsightedly stable vote allocations is a subset of the set of stable vote allocations. Nonetheless, the farsighted core is nonempty, by the same argument used to prove Proposition 1 (i.e., dictatorial vote allocations are farsightedly stable).

[24] An example of such a farsighted strategy appears in the proof of one of the propositions later in this section.

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

Table 5: *A farsightedly stable allocation relative to v_0 need not exist. An example.*

Proposition 5 *An F-stable vote allocation v exists for all z, N, and K.*

As in our previous discussion, however, what we want to know is whether F-stable vote allocations are *reachable from* v_0 via an F-chain. The definition of farsighted stability does not take into account the initial starting point. But domination chains provide the necessary dynamic link–they are the farsighted parallel to the myopic Pivot algorithm. We will call $\mathcal{V}_F(v_0)$ the set of *farsightedly stable vote allocations relative to the initial allocation* v_0: $v \in \mathcal{V}_F(v_0)$ if either v is reachable from v_0 by an F-chain and is not F-dominated, or v_0 is undominated and $v = v_0$.[25] Formally:

Definition 12 $v \in \mathcal{V}_F(v_0)$ *and thus is* farsightedly stable relative to v_0 *(an F_0-stable vote allocation) if and only if one of the following holds: either (1) $v \in D(v_0) \cap C_F$, or (2) $D(v_0) = \varnothing$ and $v = v_0$.*

Is the set $\mathcal{V}_F(v_0)$ always nonempty? Unfortunately, this is not guaranteed.

Theorem 3 *There exist N, K, and Z such that no vote allocation is farsightedly stable relative to v_0.*

We prove the theorem in Appendix A, studying the example in Table 5.

The logic is conveyed easily. In this example, v_0 cannot be F_0-stable because there exists v' that dominates v_0. At v_0, $\mathbf{P}(v_0) = \{B\}$. But there exists a one-trade F-chain to v' such that $\mathbf{P}(v') = \{A\}$: voter 1 can give a B vote to 3 in exchange for an A vote, and the trade is profitable for both. Allocation v' F-dominates v_0, but v' is not stable either: again there exists a one-trade F-chain to v'' such that $\mathbf{P}(v'') = \{B\}$: voter 2 gives a B vote to 4 in exchange for an A vote, and again the

[25] As in the case of the Pivot algorithm, at v_0 multiple F-chains might exist. Because this definition considers all possible F-chains, we do not need to be specific about the selection rule, R, along any particular chain.

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

Table 6: *A myopic coalitional trade cannot replicate a farsighted chain. An example.*

trade is profitable for both. Thus $v' \notin \mathcal{V}_F(v_0)$. Note that v'' is not reachable via an F-chain from v_0; thus $v'' \notin D(v_0)$.[26] The proof in Appendix A shows that v' is the unique vote allocation that F-dominates v_0. It then follows that $\mathcal{V}_F(v_0)$ is empty.

The trades we just described are one-step F-chains of pairwise trades. Yet, extending the analysis to farsightedness undoes our earlier stability result for pairwise trade. Whereas pairs of traders exchanging votes myopically are guaranteed to reach a vote allocation from which further myopic trades cannot be profitable, this is no longer the case with farsighted trading. In our example, the vote allocation v'' is both myopically and farsightedly stable (it belongs in the farsighted core). It is the unique Pivot-stable allocation, and would be reached by myopic traders in two steps.[27] *But $v''$is not reachable from v_0 by farsighted traders* because it does not F-dominate v_0. Convergence to a stable allocation breaks down.

The link between farsighted and myopic trades deserves more comment. Because a farsighted chain is driven by the comparison between the final vote allocation and the vote allocation at which each trade occurs, it clearly differs from a sequence of myopic trades, as the example shows. More surprising is that it may be impossible to replicate a farsighted chain with a myopic coalitional trade when the coalition can assume any arbitrary coalition size. We find:

Proposition 6 *There exist N, K, Z, and v such that v is farsightedly stable relative to v_0, but is not Pivot-stable, for any trading coalition size.*

Proof. Consider the example in Table 6.

Call $\mathcal{V}_M(v_0)$ the set of vote allocations reachable from v_0 via myopic coalitional trade. Note that $\mathcal{V}_M(v_0) = \{v_0\}$: no trade can move the vote allocation away from

[26]Both F-chains are one-trade F-chains, or equivalently myopic payoff improving trades. But recall that an allocation cannot be F-undominated if it is myopically dominated.

[27]In this example, v'' is the unique Pivot-stable allocation whether or not coalitions are restricted to be pairwise.

v_0 and myopically benefit all coalition members, for any coalition size. Thus v_0 is the unique Pivot-stable vote allocation for any coalition size. However, there exists a vote allocation $v \neq v_0$ that belongs to the F-core and F-dominates v_0. Consider the following two-step F-chain ending at v, with $\mathbf{P}(v) = \{B\}$: starting at v_0 (with $\mathbf{P}(v_0) = \{A, B\}$) 4 gives a B vote to 5, and 5 gives an A vote to 4 (and thus $\mathbf{P}(v_1) = \{A\}$); at v_1, 3, 4, and 5 transfer all their votes to 4, reaching v such that $\mathbf{P}(v) = \{B\}$ and making 4 dictator. Because 4 is dictator and obtains his preferred outcome, v is in the F-core. Hence $v \in \mathcal{V}_F(v_0)$: v is farsightedly stable relative to v_0. And yet from v_0 it cannot be reached by any myopic coalitional trade.$\square$

The logic behind the two examples in Tables 5 and 6 is very similar. In an F-chain, every effective coalition compares the final outcome reachable through the chain to the outcome corresponding to the vote allocation in place when the coalition trades–*not to the starting vote allocation*. It is then possible for a coalition to trade at some point along the chain even though the final outcome is not preferred to the initial outcome by all coalition members.

Given Theorem 3, a natural question is whether the definition of farsighted stability should be weakened to guarantee existence. Different farsighted stability concepts have been proposed in the literature–most noticeably the Bargaining set (Maschler 1992) and farsighted extensions of the von Neumann and Morgenstern stable set (Harsanyi 1974, Chwe 1994, Dutta and Vohra 2015, Ray and Vohra 2015)–to overcome the problem of an empty F-core. We discuss these alternative approaches to farsighted vote trading in Appendix B.[28] But note that our problem is different: we know that the F-core is not empty; the question is whether a domination chain can reach the F-core, *starting from* v_0. We also know, from the example in Table 6, that for some parameter values the answer is positive: $\mathcal{V}_F(v_0)$ is not empty, and F_0-stable allocations do exist.

3.2 Preferences over farsightedly-stable outcomes

Call outcome $\mathbf{P}(v)$ farsightedly-stable relative to v_0 (F_0-stable) if $v \in \mathcal{V}_F(v_0)$. If F_0-stable outcomes exist, do they possess desirable qualities? More precisely, in line

[28]Briefly: we can use the logic of the Bargaining Set, to guarantee that $\mathcal{V}_F(v_0)$ is not empty. Building on the F-stable set is more ambitious and more difficult. In particular, existence is not guaranteed.

with our analysis of myopic trade, we ask two questions: (1) When they exist, are F_0-stable outcomes always in the Pareto set? (2) Is there a relationship between F_0-stable outcomes and the Condorcet winner?

The answer to the first question follows immediately from the myopic analysis. Recall that an allocation not in the Pareto set is not in the myopic core, and hence is not in the F-core. Thus Theorem 2 extends directly to farsightedness:

Theorem 4 *If $v \in \mathcal{V}_F(v_0)$, then $\mathbf{P}(v)$ is Pareto optimal, for all K, N, and z.*

The second question–whether there is any relationship between F_0-stable outcomes and the Condorcet winner–requires a preliminary result:

Proposition 7 *If the Condorcet winner is an F_0-stable outcome, then: (1) the Condorcet winner is the only F_0-stable outcome; (2) the set of F_0-stable vote allocations $\mathcal{V}_F(v_0)$ is a singleton, and $\mathcal{V}_F(v_0) = \{v_0\}$.*

Proof. Call $\mathcal{V}_{CW}$ the set of Condorcet vote allocations. That is, $v \in \mathcal{V}_{CW} \Rightarrow \mathbf{P}(v)$ is the Condorcet winner. Suppose $v \in \mathcal{V}_{CW} \cap \mathcal{V}_F(v_0)$. Notice that if $\mathbf{P}(v) = \mathbf{P}(v_0)$, then $v \notin D(v_0)$. By Lemma 2, $v_0 \in \mathcal{V}_{CW}$. Thus if $v \in \mathcal{V}_{CW} \cap \mathcal{V}_F(v_0)$, it follows that $v = v_0$, and $v_0 \in \mathcal{V}_F(v_0)$ because no voter is better off at v than at v_0. Suppose there exists some other $v' \notin \mathcal{V}_{CW}$ such that $v' \in \mathcal{V}_F(v_0)$. Then $v' \in D(v_0)$. But then, there exists a vote allocation, v' that dominates v_0 and is not dominated by any other, since $v' \in \mathcal{V}_F(v_0)$. Hence $v_0 \notin \mathcal{V}_F(v_0)$, a contradiction. Hence it follows that $\mathcal{V}_F(v_0) = \{v_0\}$.□

In words: We know from Lemma 2 that if the Condorcet winner exists it must equal $\mathbf{P}(v_0)$. It then follows immediately that no other vote allocation yielding the Condorcet winner can farsightedly dominate v_0, and thus if an F_0-stable allocation yielding the Condorcet winner exists, it must equal v_0. But if v_0 is F_0-stable, no other allocation reachable from v_0 can be F_0-stable, because it would have to dominate v_0, and thus v_0 could not be F_0-stable. It follows that the set of F_0-stable equilibria must be a singleton and equal v_0.

Note the immediate Corollary:

Corollary 2 *The Condorcet winner can be an F_0–stable outcome only if no vote trading takes place.*

The Corollary is different from a related a result in Park (1967). Park studied vote trading when trades are revocable promises. Because the majority can always deviate jointly and implement its preferred outcome, he concluded that the only stable vote allocation is the initial pre-trade allocation when the Condorcet winner exists, and thus such initial allocation delivers it.[29] Park did not specify the process of renegotiation, nor did he specify whether traders are myopic or farsighted, and in the latter case did not offer a formal definition of farsightedness, but the conclusion that the Condorcet winner is stable because trades are revocable must implicitly embody some forward-looking logic. We, on the other hand, study irrevocable trades under a precise definition of farsightedness and find that if the Condorcet winner is farsightedly stable, relative to v_0, then no vote trading can take place. According to Park, if the Condorcet winner exists, it is a stable outcome, and the only stable outcome. In our model, if the Condorcet winner exists *and* is stable, then it is the only stable outcome.

As a normative evaluation of vote trading, the result is rather damning. In our model, vote trading and farsightedness are incompatible with achieving the Condorcet winner: if voting yields the Condorcet winner and voters are farsighted, then it must be that no vote trading has taken place.

But can we say if and when voting will in fact yield the Condorcet winner? One result is immediate:

Proposition 8 *If $N = 3$, then for any K and Z such that the Condorcet winner exists, v_0 is the unique F_0-stable vote allocation and the Condorcet winner is the unique F_0-stable outcome.*

Proof. With $N = 3$, two voters constitute a majority. If the Condorcet winner exists, by Lemma 2 $v_0 \in \mathcal{V}_{CW}$. Thus there is no effective coalition who prefers any other outcome to $\mathbf{P}(v_0)$. Hence $v_0 \in C_F(v_0)$, and thus $\mathcal{V}_F(v_0) = \{v_0\}$: there is no trade and the Condorcet winner is the unique F_0-stable outcome.□

With $N = 3$ and a Condorcet winner, no trade can be payoff-improving by assumption. The positive result in the proposition replicates the equivalent result with myopia, and in both cases results from the lack of vote trading. More interesting,

[29] Theorems 1 and 3 in Park (1967).

and less clear, are the normative properties of vote trading when profitable trades may in principle take place. Under myopia, we know that when voting concerns two proposals only, vote trading always yields the Condorcet winner when the Condorcet winner exists. Does this result extend to farsighted trading? The answer is negative:

Proposition 9 *Suppose $K = 2$ and the Condorcet winner exists. Then there exist N and Z such that $\mathcal{V}_F(v_0)$ is not empty but contains no vote allocation yielding the Condorcet winner.*

Proof. We can prove the statement by considering again the example in Table 6. Note that $\mathbf{P}(v_0)$ is the Condorcet winner (and indeed the only Pivot-stable outcome under myopia). But as we showed above, there exists $v \in \mathcal{V}_F(v_0)$ such that $\mathbf{P}(v) = \{B\}$. By Proposition 7, if there exists $v \in \mathcal{V}_F(v_0)$ such that $v \notin \mathcal{V}_{CW}$, then $\mathcal{V}_F(v_0) \cap \mathcal{V}_{CW} = \varnothing$: there exists no F_0-stable allocation whose outcome is the Condorcet winner.$\square$

Intuitively, the result is surprising: it would seem that farsightedness should favor the achievement of the Condorcet winner, exactly as Park argued. This is not what we find: even in the one limited case in which vote trading is guaranteed to deliver the Condorcet winner under myopia, the result breaks down under farsightedness. With other configurations of voter preferences, the conclusion may be reversed, and farsightedness may deliver the Condorcet winner when myopia does not. But the proposition demonstrates that the Condorcet principle is not in general respected with farsighted vote trading.

4 Conclusions

This paper proposes a general theoretical framework for studying vote trading in committees. The framework has 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 rational vote trading algorithms that define moves out of unstable vote allocations, and thus sequences of vote allocations. The model is abstract and streamlined, with 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, the final vote takes place proposal by proposal. The model would be quite different if trading concerned bundles of proposals, as opposed to individual votes, and stability were defined directly on outcomes. Finally, each vote trade is a transfer of property right among the trading parties. The easiest image is that of an actual transfer of physical ballots–thus trades cannot be reversed unilaterally (they can be reversed if all parties agree), and votes can be freely re-traded to others in future trades.

We find that vote trading need not produce vote allocations and outcomes with desirable properties. The achievement of stability is guaranteed only if vote trading is myopic and restricted to be pairwise. But the Pareto optimality of any stable vote allocation reached via trading, if such an allocation exists, is guaranteed only if vote trading is neither myopic nor pairwise. In general, there is no logical relation between the existence of a Condorcet winner and the existence and properties of a stable vote allocation. There are special cases–the most interesting being the case of two proposals only–where myopic vote trading always reaches the Condorcet winner, and only the Condorcet winner, if it exists. But the result does not extend to farsighted vote trading. In fact, active vote trading, farsightedness, and the achievement of the Condorcet winner are incompatible: if voters are farsighted, the Condorcet winner can be the voting outcome only if the pre-trade vote allocation is stable and no vote trading takes place.

Our approach suggests a number of new directions to pursue, with a broad goal of making the model less abstract and closer to actual committees and legislatures. First, many committees operate under voting rules other than simple majority rule. Our framework can be easily modified to allow for general voting rules and arbitrary specifications of decisive coalitions, for example imposing supermajority requirement rules, or empowering some agents or coalitions with veto power.

A second, related extension would be to consider different restrictions on the coalitions that can organize vote trades. We have allowed for two possibilities only: pairwise trading (i.e., any coalition of exactly two voters can enforce a trade); and unlimited coalitional trading (i.e., no restriction at all on the coalitions that can organize a trade). But in some committees or legislatures, norms or party ties may limit which coalitions can form. For example, in some cases it is difficult to engage vote trades that cross ideological or party lines. In other cases, the leadership within the committee may play a key role in negotiating and enforcing agreements. There is a large set of possibilities in this vein that would be interesting to explore.

A third direction concerns agenda setting and agenda manipulation. In the model studied in this paper, the set of binary issues is assumed to be exogenously given. In practice, the proposals up for vote are typically the outcome of an agenda formation process. There are different ways to introduce such a process into the model. In one such approach, an agenda setter or committee chair has it within his or her power to bundle proposals, and may benefit, or the committee may benefit, from doing so. Alternatively, the committee may hold a vote over how to bundle a large number of proposals into a smaller number of proposals. Our results suggest that reducing the number of effective proposals–reducing the number of possible trades–may in some circumstances be beneficial. In the extreme case where all proposals are bundled into a single omnibus proposal (not a rare occurrence in some committees and legislatures) all possible vote trades are eliminated. Agenda setting in the form of bundling introduces a new, interesting perspective on modeling logrolling in committees.

Finally, a different but important question is how to incorporate uncertainty in the model. Our framework has no formal inclusion of uncertainty and the preferences of voters are essentially considered as common knowledge.[30] 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 subjects, 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. Subjects seem sensitive to the strategic uncertainty they face: it is difficult to predict future vote trades that might be triggered by a current vote trade. This Knightian-type uncertainty suggests that ambiguity aversion may play a role in the voters' behavior. More traditional modeling of uncertainty could also be incorporated, such as private information about one's own preferences, or uncertainty about an unknown state of the world that affects everyone's preferences, as in Condorcet jury models.

[30]We qualify our statement–"formal" inclusion, "essentially" common knowledge–because our dynamic rule is an algorithm.

T_{AB}	$T_\varnothing$	T_{A-}	T_{B-}	T_{A+}	T_{B+}
A, B	$\varnothing$	A	B	A	B
A or B	A or B	$\varnothing$	$\varnothing$	A, B	A, B
B or A	B or A	A, B	A, B	$\varnothing$	$\varnothing$
$\varnothing$	A, B	B	A	B	A

Table 7: $K = 2$, *possible preference types.*

Appendix A. Proofs

Proposition 3 *If $K = 2$, then, for all N, z, and R: (1) (a) $\mathcal{V}_R^*$ is nonempty and (b) $\mathcal{P}(\mathcal{V}_R^*)$ is a singleton. (2) $\mathbf{P}(\mathcal{V}_R^*)$ is Pareto optimal. (3) If a Condorcet winner exists, then $\mathbf{P}(\mathcal{V}_R^*)$ is the Condorcet winner. (4) If $\mathbf{P}(\mathcal{V}_R^*) \neq \mathbf{P}(v_0)$, then a majority prefers $\mathbf{P}(\mathcal{V}_R^*)$ to $\mathbf{P}(v_0)$.*

Proof of (1a): *$\mathcal{V}_R^*$ is nonempty.*

Without loss of generality, we assume throughout the proof that $\mathbf{P}(v_0) = \{A, B\}$. We begin with some preliminary observations.

First, by Lemma 1, at any even step t, a Pivot algorithm can only go from $\mathbf{P}(v_t) = \{A, B\}$ to $\mathbf{P}(v_{t+1}) = \varnothing$, because a trade must cause two changes in outcomes in order for the trading voters to all benefit from the trade. Similarly, at any odd step t, a Pivot algorithm can only go from $\mathbf{P}(v_t) = \varnothing$ to $\mathbf{P}(v_{t+1}) = \{A, B\}$.

Second, because preferences over proposals are additively separable, there are exactly eight possible voter preference orders. These are summarized in Table 7, which combines the two preference types that prefer both proposals to pass, and combines the two preference types that prefer both proposals to fail (for reasons that will be clear below). An outcome in any cell is strictly preferred by that preference type to all outcomes below it.

Third, at each even step of a pivot trading sequence $t = 0, 2, ..., \mathbf{P}(v_t) = \{A, B\}$ and any coalition that blocks v_t must include a T_{A-} voter and a T_{B-} voter. To see this, note that in order to switch proposal A from pass to fail, some voter must trade away a "yes" vote on proposal A, so that voter must prefer $\{A\}$ to $\varnothing$. That voter must also benefit from the trade, which implies that the voter also must prefer $\varnothing$ to $\{B\}$ and prefer $\varnothing$ to $\{A, B\}$, and hence is a T_{A-} voter. Similarly, in order to switch proposal B from pass to fail, some voter must trade away a "yes" vote on proposal B, so that voter must prefer $\{B\}$ to $\varnothing$. That voter must also benefit from the trade, which implies that the voter also must prefer $\varnothing$ to $\{A\}$ and prefer $\varnothing$ to $\{A, B\}$, and hence is a T_{B-} voter. Following a similar logic, at each odd step of a Pivot trading sequence $t = 1, 3, ...$, $\mathbf{P}(v_t) = \varnothing$ and any coalition that blocks v_t must include a T_{A+} voter and a T_{B+} voter. To see this, note that in order to switch proposal A from fail to pass, some voter must trade away a "no" vote on proposal A, so that voter must prefer $\varnothing$ to $\{A\}$. That voter must also benefit from the trade, which implies that the voter also must prefer $\{B\}$ to $\varnothing$ and prefer $\{A, B\}$ to $\varnothing$, and hence is a T_{B+} voter. Similarly, in order to switch proposal B from fail to pass, some voter must trade away a "no" vote on proposal B, so that voter must prefer $\varnothing$ to $\{B\}$. That voter must also benefit from the trade, which implies that the voter also must prefer $\{A\}$ to $\varnothing$ and prefer $\{A, B\}$ to $\varnothing$, and hence is a T_{A+} voter.

Fourth, T_{AB} voters and $T_{\varnothing}$ voters are never part of a trading coalition. Consider even t, so that $\mathbf{P}(v_t) = \{A, B\}$. (The logic is similar for odd t.) Clearly T_{AB}, T_{A+}, and T_{B+} voter types cannot be trading because they strictly lose by going from $\{A, B\}$ to $\varnothing$. Hence, it is only T_{A-} voters who trade away "yes" votes on proposal A and only T_{B-} voters who trade away "yes" votes on proposal B. Nobody can be trading away a "no" vote on either proposal, as these would be redundant trades and violate minimality. Hence the only possibility remaining is that some $T_{\varnothing}$ voter receives a vote without trading any votes away. But such a voter is not needed, because any votes she received on proposal A could have been traded instead to one of the T_{B-} voters in the coalition and any votes she received on proposal B could have been traded instead to one of the T_{A-} voters in the coalition. Hence coalition minimality implies that the coalition can be populated only by T_{A-} and T_{B-} voters.

The proof now proceeds by a series of simple lemmas.

Lemma 3 *If $K = 2$ and at v_t both proposals are decided by minimal majority, then all Pivot trades at all $t' \geq t$ must be pairwise trades.*

Proof. By minimality of the trades, if at v_t both proposals pass by minimal majority then at each vote allocation $v_{t'}$, $t' > t$, both proposals must be decided by minimal majority. Assume t even. From above, any blocking coalition at v_t must include a T_{A-} voter trading away a "yes" vote on proposal A and a T_{B-}

voter trading away a ”yes” vote on proposal B. But if the proposals are decided by minimal majority, those two voters could just trade with each other. Hence the minimal coalition must be a pair.□

Lemma 4 *If $K = 2$, there can be at most one Pivot trade that is not pairwise, and it can only be the first.*

Proof. By Lemma 3, if non-pairwise 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. Hence, by Lemma 3, from $t = 1$ onward, all trades can only be pairwise trades.□

If v_0 cannot be blocked, then it is trivially stable. If it can be blocked, then by Lemmas 3 and 4, for all $t > 0$, v_t has the property that both proposals are decided by a minimal majority and all future minimal blocking coalitions are pairs. Since they are pairs, scores monotonically increase after the first period, and hence trading must end at some point. Therefore $\mathcal{V}_R^*$ is always nonempty, so (1a) is proved.□

Proof of (1b): *$\mathcal{V}_R^*$ is unique.*

The following intermediate result is useful.

Lemma 5 *If $K = 2$, each voter takes part in at most one trade.*

Proof. We show that T_{A-} voters only trade once. The logic is the same for T_{B-}, T_{A+}, and T_{B+} voter types. The key observation here is that if a T_{A-} voter, call her i, is in a trading coalition she must be trading away a ”yes” vote on proposal A. If not, then she is receiving some proposal B votes and there is at least one other T_{A-} voter in the coalition, call her j, who trades away a ”yes” vote on A. Any proposal B votes that i received could have been traded to j instead, and so i is a redundant member of the coalition. Hence the coalition was not minimal. Since each T_{A-} voter starts with a single vote on each issue, then as soon as such a voter trades, he or she holds no remaining proposal A votes and hence will never trade again.□

We next show that at each step, if there are multiple minimal blocking trades they are all equivalent in the sense that they all result in the same future trading possibilities. Consider $t = 0$. If there is no blocking coalition, v_0 is stable, $\mathcal{V}_R^*$ is unique, and (1b) holds. Suppose instead that there is at least one blocking coalition. At v_0, proposal A passes with a margin of victory of m_A votes and proposal B with a margin of m_B votes (for example, $m_A = m_B = 1$ when both proposals pass by minimal majority). Then, from the earlier argument in the proof of (1a), any minimal blocking coalition must consist of exactly m_A T_{A-} voters and exactly m_B T_{B-} voters. There may be multiple such coalitions and within each such coalition the trade may result in different vote allocations. But all subsequent vote allocations, v_1, will be equivalent in the sense that they all imply the same set of possible future trading opportunities. This is true because the executed trades only differ in the labels of the voters in the coalitions, not in the profile of voter types. Hence the profile of voter types and voter allocations in the complement of the coalition are the same. Since, by Lemma 5, voters in the trading coalition are unable to engage in any future trades, the future trading opportunities only depend on the profile of voter types and voter allocations in the complement of the coalition. The same logic applies to all future even steps, with the further restriction (by Lemma 4) that $m_A = m_B = 1$. A similar argument applies to odd steps, except that all trading pairs have to consist of exactly one T_{A+} voter and exactly one T_{B+} voter (again by Lemma 4). Hence all trading paths are essentially equivalent and trading will stop either at the first even step where there does not exist a matched pair of T_{A-} and T_{B-} voters who have not yet traded (in which case $\mathcal{P}(\mathcal{V}_R^*) = \{A, B\}$), or the first odd step where there does not exist a matched pair of T_{A+} and T_{B+} voters who have not yet traded (in which case $\mathcal{P}(\mathcal{V}_R^*) = \varnothing$), whichever occurs first.□

Proof of (2): $\mathbf{P}(\mathcal{V}_R^*)$ *is Pareto optimal.*

In $\mathbf{P}(\mathcal{V}_R^*)$, the outcome on each proposal (either pass or fail) is preferred by at least $\frac{n+1}{2}$ voters. Since there are only two proposals, there exists at least one voter who is on the winning side of both proposals. For this voter $\mathbf{P}(\mathcal{V}_R^*)$ is the most preferred outcome, and hence $\mathbf{P}(\mathcal{V}_R^*)$ is Pareto optimal. Note that the same logic of the proof applies to any vote allocation, v. That is, with $K = 2$, $\mathbf{P}(v)$ is Pareto optimal for every vote allocation $v \in \mathcal{V}$. □

Proof of (3): *If a Condorcet winner exists, then* $\mathbf{P}(\mathcal{V}_R^*)$ *is the Condorcet winner.*

Denote by n_{AB} the number of voters of type T_{AB}, and similarly for the other types defined in Table 7. By Lemma 2, if the Condorcet winner exists it can only be $\mathbf{P}(v_0)$. Suppose then, WLOG, that $\mathbf{P}(v_0) = \{A, B\}$ is the Condorcet winner. Denote by m_{AB} the (positive) difference between the number of voters who prefer $\{A, B\}$ to $\varnothing$, and those who prefer $\varnothing$ to $\{A, B\}$. At v_0, proposal A passes with a margin of m_A votes, and proposal B passes with a margin of m_B votes, where m_A, m_B, and m_{AB} are all positive integers. Thus we must have, respectively:

$$\begin{aligned} n_{AB} + n_{A+} + n_{B+} &= \frac{n-1}{2} + m_{AB} \\ n_{AB} + n_{A-} + n_{A+} &= \frac{n-1}{2} + m_A \\ n_{AB} + n_{B-} + n_{B+} &= \frac{n-1}{2} + m_B \end{aligned}$$

If there is no blocking coalition at v_0, v_0 is Pivot stable, and $\mathbf{P}(\mathcal{V}_R^*)$ $=\mathbf{P}(v_0)$ is the Condorcet winner. Suppose instead that there is a blocking coalition at v_0. From the arguments above, any minimal blocking coalitions must consist of exactly m_A voters of type T_{A-} and exactly m_B voters of type T_{B-}. We now show that this implies $n_{A+} > 0$ and $n_{B+} > 0$, and hence there is a pairwise coalition that blocks v_1, where $\mathbf{P}(v_1) = \varnothing$. Suppose to the contrary that $n_{A+} = 0$. From the second equation, this implies $n_{AB} \leq \frac{n-1}{2}$ because $n_{A-} \geq m_A$, or there could be no blocking coalition at v_0. Plugging $n_{AB} \leq \frac{n-1}{2}$ into the first equation gives us $n_{A+} + n_{B+} \geq m_{AB}$, but $n_{A+} = 0$ by supposition, so $n_{B+} \geq m_{AB} > 0$. But then from the third equation, $n_{B+} > 0$ implies that $n_{AB} \leq \frac{n-3}{2}$ because $n_{B-} \geq m_B$, or there could be no blocking coalition at v_0. Returning to the first equation, this implies $n_{B+} > 1$, which (from the third equation) implies $n_{AB} \leq \frac{n-5}{2}$, and so forth. But since $\mathbf{P}(v_0) = \{A, B\}$, $n_{AB} \geq 1$. Hence eventually we reach a contradiction. Thus $n_{A+} > 0$. By a similar argument, we obtain $n_{B+} > 0$, so there exists a pairwise coalition that blocks v_1. Therefore v_1 is not Pivot-stable and $\mathbf{P}(v_2) = \{A, B\}$. Note that at v_2, A and B both pass by exactly one vote.

The following three equations must hold:

$$
\begin{aligned}
n_{AB} + n_{A+} + n_{B+} &= \frac{n-1}{2} + m_{AB} \\
n_{AB} + (n_{A-} - m_A) + (n_{A+} + 1) &= \frac{n+1}{2} \\
n_{AB} + (n_{B-} - m_B) + (n_{B+} + 1) &= \frac{n+1}{2}
\end{aligned}
$$

As before, the first equation reports preferences rankings between $\{A, B\}$ and $\varnothing$ and is unchanged. The second equation mirrors the one-vote advantage in favor of A. Relative to v_0, the number of votes cast in favor of A is reduced by the m_A voters of type n_{A-} who traded their "yes" vote on A away at v_0 and increased by the single n_{A+} voter who received an extra vote to cast as a "yes" vote on A at v_1. The third equation mirrors the one-vote advantage in favor of B, and shows the equivalent changes with respect to v_0.

If there is no blocking coalition at v_2, then v_2 is Pivot stable, and $\mathbf{P}(\mathcal{V}_R^*) = \mathbf{P}(v_2) = \{A, B\}$ is the Condorcet winner. Suppose instead that there is a blocking coalition at v_2. Such coalition must consist of exactly one voter of type T_{A-} and exactly one voter of type T_{B-}, neither of whom has traded yet.

We now show that if a blocking coalition exists at v_2, then the three equations imply $n_{A+} > 1$ and $n_{B+} > 1$, i.e. there must exist at least one voter of type T_{A+} and at least one voter of type T_{B+} who have not traded yet, and thus if there is a blocking coalition at v_2, then there must be a blocking coalition at v_3, where $\mathbf{P}(v_3) = \varnothing$.

Suppose to the contrary that $n_{A+} = 1$. From the second equation, this implies $n_{AB} \leq \frac{n-5}{2}$ because $(n_{A-} - m_A) \geq 1$, or there would not be a blocking coalition at v_2. Plugging $n_{AB} \leq \frac{n-5}{2}$ into the first equation gives us $n_{A+} + n_{B+} \geq m_{AB} + 2$, but $n_{A+} = 1$ by supposition, so $n_{B+} \geq m_{AB} > 1$. But then from the third equation, $n_{B+} > 1$ implies $n_{AB} \leq \frac{n-7}{2}$ because $(n_{B-} - m_B) \geq 1$, or there would not be a blocking coalition v_2. Returning to the first equation, this implies $n_{B+} > 2$, which (from the third equation) implies $n_{AB} \leq \frac{n-9}{2}$, and so forth. But $n_{AB} \geq 1$, hence eventually we reach a contradiction. Thus $n_{A+} > 1$. By a similar argument, we obtain $n_{B+} > 1$. Therefore there exists a pairwise coalition that blocks v_3. Therefore $\mathbf{P}(v_4) = \{A, B\}$. A similar logic applies to any subsequent trading steps. If there is a blocking coalition that moves the outcome to $\varnothing$, then there is another blocking

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

Table 8: *The set of Pivot-stable outcomes can be a superset of the Uncovered set. An example.*

coalition at the next step that moves the outcome back to $\{A, B\}$. Therefore, since there are at most a finite number of steps, trading can only stop at some terminal v_T, with $\mathbf{P}(v_T) = \{A, B\}$.□

Proof of (4): *If* $\mathbf{P}(\mathcal{V}_R^*) \neq \mathbf{P}(v_0)$*, then a majority prefers* $\mathbf{P}(\mathcal{V}_R^*)$ *to* $\mathbf{P}(v_0)$*.*

If $\mathbf{P}(v_0)$ is the Condorcet winner, then, by result (3), $\mathbf{P}(\mathcal{V}^*) = \mathbf{P}(v_0)$. Thus if $\mathbf{P}(\mathcal{V}^*) \neq \mathbf{P}(v_0)$, then $\mathbf{P}(v_0)$ is not the Condorcet winner. Suppose $\mathbf{P}(v_0) = \{A, B\}$. Then $\mathbf{P}(v_0)$ is majority-preferred to both $\{A\}$, and $\{B\}$. But if $\mathbf{P}(v_0)$ is not the Condorcet winner, it must be that $\varnothing$ is majority-preferred to $\{A, B\}$. If $\mathbf{P}(\mathcal{V}^*) \neq \mathbf{P}(v_0)$, then $\mathbf{P}(\mathcal{V}^*) = \varnothing$, and thus $\mathbf{P}(\mathcal{V}^*)$ is majority-preferred to $\mathbf{P}(v_0)$.□

The set of Pivot-stable outcomes can be a superset of the Uncovered set.

Consider the example in Table 8.

The Uncovered set consists of $\{A, B, C\}$, the Condorcet winner. But with these preferences, $\mathbf{P}(\mathcal{V}_R^*) = \{\{A, B, C\}, \{A\}, \{C\}\}$. The corresponding sets of pairwise trades are: $\{23ba, 34ca, 45cb\}$ leading to $\{A, B, C\}$, $\{23ba, 45ab, 13bc\}$ leading to $\{A\}$, and $\{13bc, 45cb, 23ba\}$ leading to $\{C\}$.

Theorem 5 *There exist* N*,* K*, and* Z *such that no vote allocation is farsightedly stable relative to* v_0*.*

Proof:

Consider the example in Table 9. The lower panel reports, in the column corresponding to each voter, the voter's ordinal preferences over the four possible outcomes. An outcome in a cell is strictly preferred by that voter to all outcomes in lower cells.

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

AB	$\varnothing$	$\varnothing$	AB	B
A	B	A	B	AB
B	A	B	A	$\varnothing$
$\varnothing$	AB	AB	$\varnothing$	A

Table 9: *No vote allocation is farsightedly stable relative to* v_0*. An example.*

Note that $\mathbf{P}(v_0) = \{B\}$. The proof is in two steps. We first show that there is a unique vote allocation v that F-dominates v_0, and v must be such that $\mathbf{P}(v) = \{A\}$. We then show that $v \notin C_F$.

(1). By definition of $D(v_0)$, any $v \in D(v_0)$ must be such that $\mathbf{P}(v) \neq B$. (i) Suppose there exists some $v' \in D(v_0)$ such that $\mathbf{P}(v') = \{A, B\}$. However, $\{A, B\}$ is the least preferred alternative for voters 2 and 3, so those two voters never trade as part of an F-chain to v'. Voter 5 ranks $\mathbf{P}(v_0) = \{B\}$ above $\{A, B\}$; hence will not trade at v_0. Therefore, at v_0, on an F-chain to v' such that $\mathbf{P}(v') = \{A, B\}$, the only possible first trade is between voters 1 and 4. But 1 and 4 have identical preferences, and no trade between them can change the outcome. Hence no trade between them can advance the F-chain: there cannot exist a $v' \in D(v_0)$ such that $\mathbf{P}(v') = \{A, B\}$. (ii) Similarly, there cannot exist a $v' \in D(v_0)$ such that $\mathbf{P}(v') = \varnothing$: the voters' preference rankings are such that only voters 2 and 3 can trade at v_0 on an F-chain to any v' such that $\mathbf{P}(v') = \varnothing$. But 2 and 3 have identical preferences and cannot advance the F-chain. (iii) Is there some $v' \in D(v_0)$ such that $\mathbf{P}(v') = \{A\}$? Voter 5 never trades on such an F-chain. Therefore, at v_0, on an F-chain to v' such that $\mathbf{P}(v') = \{A\}$, the only possible first trade is between voters 1 and 3. They disagree on both proposals, and thus by trading can reach any outcome. They can trade to any of the eight vote allocations shown below (where the number in each cell indicates the number of votes 1 and 3 hold after the trade; all other voters hold one vote):

	1	3	1	3	1	3	1	3	1	3	1	3	1	3	1	3
A	1	1	0	2	0	2	2	0	2	0	0	2	1	1	2	0
B	2	0	1	1	2	0	2	0	1	1	0	2	0	2	0	2

The first three of these vote allocations do not change the outcome, and thus cannot advance the F-chain.

The next two change the outcome to $\{A, B\}$. If either of these trades occur at v_0, then the only subsequent trade from either of these two vote allocations on an F-chain to v' such that $\mathbf{P}(v') = \{A\}$, can only have 2 and 3 trading (because they are the only two voters who prefer $\{A\}$ to $\{A, B\}$; but 2 and 3 both have negative values for both proposals and therefore cannot advance the F-chain.

The next two possible vote trades between 1 and 3 will lead to the $\varnothing$ outcome. On an F-chain to v' such that $\mathbf{P}(v') = \{A\}$ only 1 and 4 can gain from moving from $\varnothing$ to $\{A\}$; but 1 and 4 both have positive values for both proposals and therefore cannot advance the F-chain.

Finally, consider the last possible vote allocation, which results when 1 trades his B vote for 3's A vote, leading to the outcome $\{A\}$. Call this vote allocation v', since $v' \in D(v_0)$ and $\mathbf{P}(v') = \{A\}$. Because we have ruled out all other possible allocations, $D(v_0)$ is a singleton: $D(v_0) = \{v'\}$.

(2). But $v' \notin C_F$: voter 2 can give a B vote to voter 4, in exchange for an A vote, and reach the vote allocation v'' depicted below, with $\mathbf{P}(v'') = \{B\}$. Both 2 and 4 prefer the outcome $\{B\}$ to $\{A\}$. The reasoning shows that v' is not myopically stable, and therefore is not in the F-core.

	1	2	3	4	5
A	2	2	0	0	1
B	0	0	2	2	1

v''

We have shown that $v_0 \notin C_F$, since there exists $v' \in D(v_0)$, and that all $v \in D(v_0)$ are such that $v \notin C_F$, since $D(v_0) = \{v'\}$ and $v' \notin C_F$. Hence the set $D(v_0) \cap C_F$ is empty: no vote allocation is farsightedly stable relative to v_0. $\square$

Appendix B. Alternative notions of farsighted stability

We approach farsighted stability in the text using the forward looking extension of the core of a game without side payments. Alternative notions could be explored. We briefly present two such alternatives in this Appendix.

The F-Bargaining Set

The farsighted bargaining set weakens the notion of the farsighted core by allowing dominated allocations to belong to the set.

Formally:

Definition 13 *A vote allocation $v \in \mathcal{V}$ is in* the farsighted bargaining set*, B_F, if, for every $v' \in \mathcal{V}$ such that v' F-dominates v, there exists some $v'' \in V$ such that v'' F-dominates v'. That is, $B_F = \{v|\ D(v') \neq \varnothing\ \forall v' \in D(v)\}$.*

The farsighted bargaining set contains the core, and therefore is nonempty (by Proposition 5).

We can define the *F-bargaining set reachable from v_0, $\mathcal{V}_B(v_0)$*:

Definition 14 *A vote allocation $v \in \mathcal{V}_B(v_0)$ if and only if one of the following holds: either (1) there exists $v \in D(v_0) \cap B_F$, or (2) $D(v_0) \cap B_F = \varnothing$ and $v = v_0$.*

Proposition 10 *$v \in \mathcal{V}_B(v_0)$ is nonempty for all z, N, and K.*

Clearly $\mathcal{V}_B(v_0)$ is never empty. The result is by construction but reflects the spirit of the concept: if no vote allocation that dominates v_0 belongs to the F-bargaining set (i.e. if $D(v_0) \cap B_F = \varnothing$), then v_0 itself should be understood as belonging to F-bargaining set reachable from v_0 –although v_0 may be dominated by some other allocations, none of these allocations is itself robust to further credible domination.

Some, but not all of the results from sections 3.1 and 3.2 apply with this alternative definition. First note that the definition of $\mathcal{V}_F(v_0)$ in the text (Definition 12) ensures that $v \in \mathcal{V}_F(v_0) \implies v \in \mathcal{V}_B(v_0)$. Therefore, it immediately follows that Propositions 6 and 8 continue to hold. The other results, however, fail to extend to this alternative definition because $\mathcal{V}_B(v_0)$ can include vote allocations that are neither in the farsighted core nor equal to v_0. The definition of $\mathcal{V}_B(v_0)$ can be strengthened to rule out such possibilities by replacing $v \in D(v_0) \cap B_F$ and $v = v_0$ if $D(v_0) \cap B_F = \varnothing$ by $v \in D(v_0) \cap C_F$ and $v = v_0$ if $D(v_0) \cap C_F = \varnothing$. With this stronger definition, non-emptiness is still guaranteed, and Propositions 7 and 9 and Corollary 2 hold.

The von Neumann Morgenstern (NM) F-stable set

A second possible approach is to use a farsighted generalization of von Neumann-Morgenstern stable sets. Defining and analyzing farsighted stable sets is much more involved than analyzing the farsighted core and bargaining sets because the concept is defined as a set-valued fixed point in a space with no natural topology.

Here we extend the definition of the *NM Farsightedly Stable (NMF-Stable) Set* $\mathcal{V}_{NM}$, originally proposed by Harsanyi (1974) to our vote trading environment. Following Ray and Vohra (2015), for any subset of vote allocations, $\mathbf{V} \subseteq \mathcal{V}$, define $dom(\mathbf{V})$ as the set of vote allocations that are farsightedly dominated by some allocation $v \in \mathbf{V}$. A set of *NMF-stable* vote allocations $\mathcal{V}_{NM}$ has the property that no vote allocation in $\mathcal{V}_{NM}$ is dominated by another vote allocation in $\mathcal{V}_{NM}$ (*internal stability*) and every feasible vote allocation not in $\mathcal{V}_{NM}$ is dominated by at least one vote allocation in $\mathcal{V}_{NM}$ (*external stability*). Formally:

Definition 15 *$\mathcal{V}_{NM} \subseteq \mathcal{V}$ is an* NMF-stable set *if $\mathcal{V}_{NM} = \mathcal{V} - dom(\mathcal{V}_{NM})$.*

Note that the definition is set-based: in general, which allocations belong to the set depends on the full set itself. Neither existence nor uniqueness are guaranteed. An additional difficulty is that in practical applications verifying whether an allocation is F-stable requires positing the full composition of the set–a difficult task.[31]

As with the F-bargaining set and the F-stable set, one needs to extend the definition of NMF-stable sets to require reachability from v_0. Define $dom_{D(v_0)}(\mathcal{V}_{NM})$ the set of allocations in $D(v_0)$ that are F-dominated by some allocation in $\mathcal{V}_{NM}$.[32] Then:

[31] Which is why most progress has been made in cases in which the NMF-set can be restricted a priori to be a singleton (Mauleon et al. 2011, Ray and Vohra, 2015).

[32] Note that $dom_{D(v_0)}(\mathcal{V}_{NM})$ is defined with respect to $\mathcal{V}_{NM}$, not $\mathcal{V}_{NM}(v_0)$.

Definition 16 $\mathcal{V}_{NM}(v_0)$ *is an* NMF-stable set reachable from v_0 *if, given a set* $\mathcal{V}_{NM}$*,* $\mathcal{V}_{NM}(v_0) = D(v_0) - dom_{D(v_0)}(\mathcal{V}_{NM})$.

Unfortunately, existence of at least one such $\mathcal{V}_{NM}(v_0)$ is not guaranteed.

References

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

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

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

[4] Chwe, M., 1994, "Farsighted Coalitional Stability", *Journal of Economic Theory*, 63, 299-325.

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

[6] Diamantoudi, E. and L. Xue, 2003, "Farsighted stability in hedonic games", *Social Choice and Welfare*, 21: 39-61

[7] Diamantoudi, E., E. Miyagawa, and L. Xue, 2004, "Random Paths to Stability in the Roommate Problem", *Games and Economic Behavior*, 48, 18-28.

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

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

[10] Dutta, B. and R. Vohra (2015), "Rational Expectations and Farsighted Stability", *Working Paper*, Brown University.

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

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

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

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

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

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

[17] Harsanyi, J. 1974. "An Equilibrium-point Interpretation of Stable Sets and a Proposed Alternative Definition", *Management Science*, 20: 1472-95.

[18] Jackson, M. and A. Watts, 2002, "The Evolution of Social and Economic Networks", *Journal of Economic Theory*, 106(2): 265-295.

[19] Jackson, M. and A. Wolinsky, 1996, "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, 71(1): 44-74.

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

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

[22] Maschler, M., 1992, "The Bargaining Set, Kernel, and Nucleolus", in *The Handbook of Game Theory, Vol. 1*, R. Aumann and S. Hart eds., 591-667.

[23] Mauleon, A., V. Van Netelbosch and W. Vergote, 2011, "von Neumann-Morgenstern farsighted stable sets in two-sided matching", *Theoretical Economics*, 6, 499-521.

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

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

[26] Ray, D. and R. Vohra, 2015. "The Farsighted Stable Set", *Econometrica*, 83, 977-1011.

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

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

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

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

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

[32] Tullock, Gordon. 1959. "Problems of Majority Voting", *Journal of Political Economy*. 67: 571-79.

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

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

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

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