ON THE POSSIBILITY OF INFORMATION AGGREGATION IN LARGE ELECTIONS

PAULO BARELLI, SOURAV BHATTACHARYA AND LUCAS SIGA

Abstract. We study the problem of aggregation of private information in common value elections with two or more alternatives and with general state and signal spaces, where the winning alternative is chosen via majority rule. We show that, for a generic class of environments, there exists a sequence of equilibria that efficiently aggregates information as the population size grows to infinity. The result can be interpreted in the following way: if the signal space is rich enough relative to the state space information aggregation is generic; if the state space is rich relative to the signal space, information can be efficiently aggregated only when the information structure satisfies a strong property.

1. Introduction

In any election, decision relevant information is dispersed among the voters. It is necessary to aggregate all the individuals's privately held information to ascertain who the best candidate is. But since each voter is constrained to vote with his own private information, we have the classic question of whether the electoral mechanism correctly aggregates information: Does the electoral outcome coincide with the candidate the voters would have selected if all the private information were commonly known? We analyze this question in a setting where voters have the same underlying preference: given the aggregated information, all voters would choose the same candidate.

This question was first posed and answered in the affirmative by Condorcet (1786). If each voter votes for the correct candidate with a probability $p > \frac{1}{2}$, then individual, uncorrelated errors do not matter in the aggregate and the majority is almost surely right in a large electorate by the Law of Large Numbers. In the modern interpretation of the environment, there are two states of the world, each corresponding to one candidate being the commonly preferred choice and each voter's private information is a noisy signal about the state. Recent work (e.g., Ladha (1982), Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1997), Wit (1998), Myerson (1998), Duggan and Martinelli (2005)) has generalized this result, known popularly as Condorcet Jury Theorem (CJT), to more general assumptions on information structures, voting rules, voter rationality and so on.

Date: February 2, 2017.

This paper is partly based on an earlier version with the same title, co-authored with Tim Feddersen and Wolfgang Pesendorfer. We are deeply indebted to both of them for their inputs. We are also thankful to Hari Govindan, Bard Harstad and Roee Teper for helpful suggestions. All errors that remain are our own.

The two-state formulation, however, substantially reduces the complexity of the information aggregation problem by equating a private signal to an aggregate assessment of which candidate is better. Formally, the binary state assumption implies that there are exactly two possible distributions of private signals in a large electorate. However, the electability of candidates depend on a myriad of factors like their policy positions on different issues, their past history, party affiliation, the state of the economy, the geopolitical situation and so forth. Similarly, different voters are likely to hold information about different factors, and a private signal may favor different candidates depending on the underlying situations arising as a combination of these factors. We identify by the state variable the different possible situations that affect voter preference, and our standard for informational efficiency requires the correct candidate to be elected for every possible state. Given this yardstick, the objective of our paper is to examine how the property of information aggregation depends on the relationship between (common) preference and distribution of information in the electorate. Our main message is that the substantive interpretation of signals and states in a particular environment matters for the property of information aggregation.

In the formal sense, we extend the CJT to general state and signal spaces and consider two or more alternatives. Recall that a state in our formulation encapsulates every aspect that might affect voters's preferences; likewise, a signal is meant to encapsulate every information that a voter may acquire prior to voting. Hence, it is important to allow states and signals to be as general as possible. Our results can be interpreted the following way: If there are at least as many signals as there are states, then, generically there exists an equilibrium that aggregates information. On the other hand, if the state space is infinite and the signal space is finite, it is impossible to aggregate information except for very special situations. Thus, the property of information aggregation depends on the complexity of the information structure. Aggregation failure in our setting happens not due to perverse equilibrium inferences, but because the variations in a voter's information is not sufficiently rich to be able to distinguish between the different situations over which the preferences vary. This distinction is not due to a mere identification problem of the best alternative given the aggregate information available across the whole electorate. The distinction is a particular restriction imposed by the voting mechanism even when the aggregate information dispersed in the market is sufficient to choose, with arbitrarily high confidence, the best alternative.

For illustration, consider a single-issue setting where an incumbent a_1 competes against a challenger a_2, and suppose that each voter gets one of two signals: x or x'. First assume that they are competing on quality, and the signals x and x' are good and bad news, respectively, about a_1's relative quality, in the sense that as the quality of candidate a_1 improves, each voter is more likely to have signal x and less likely to have signal x'. In this case, we are in a setting similar to the canonical CJT, and voting will aggregate information. Next, consider a different setting where there is uncertainty about a_1's policy position on the left-right dimension, and voters prefer to vote for a_1 only if her policies are sufficiently moderate. Now suppose that an x-signal arises more relatively frequently for more left-leaning positions and an x'-signal arises more frequently for more right-leaning positions. Since the signals only tell the voter about whether the position is skewed to the left or to the right and not about how extreme the position is, information aggregation is impossible for any plurality rule.[1] While the example seems a bit stylized, such problems arise routinely in settings where more than one issue is involved. Consider a referendum on a reform proposal against a status quo. The reform has uncertain consequences on two policy dimensions, and voters prefer the proposal only if it is closer to their ideal policy compared to the status quo on this two-dimensional policy space. By the same logic as above, if voters only obtain binary signals about the consequences of the reform on each issue dimension, information aggregation fails generically.

In the formal model, there are several alternatives, a state and a signal space. Each voter obtains a signal independently from a distribution over signals conditional on the state. Each member of the electorate votes for one of the alternatives, and the alternative getting most votes wins. Additionally, we assume that a state where the electorate is indifferent is unlikely.[2] We say that an environment allows Full Information Equivalence (FIE) if there exists some *feasible* strategy profile which induces the correct outcome ex-ante almost surely as the electorate size becomes large. Borrowing an insight from McLennan (1998), we show that as long as there exists a feasible strategy profile that induces FIE, there is also an equilibrium profile that does the same. Therefore, while aggregation is possible in (some) equilibrium in an environment that allows FIE, aggregation is impossible in environments that violate FIE even if voters can commit to any arbitrary strategy profile.

Our first result (Theorem 1) provides a necessary and sufficient condition for FIE in environments with two alternatives (say, a_1 and a_2). Denote the set of conditional distributions arising in states where a_1 (resp. a_2) is preferred by $\mathcal{A}_1^\Delta$ (resp. $\mathcal{A}_2^\Delta$). An environment allows FIE if and only if $\mathcal{A}_1^\Delta$ and $\mathcal{A}_2^\Delta$ are separable by a hyperplane on the simplex. This result (and most others in the paper) follows from the observation that the expected vote share for an alternative (given any strategy) is a linear function of the vectors on the simplex, which is a direct implication of each voter having to vote only based on his own private information.

One must note here that we do not allow communication between voters in our model. Clearly, since everyone has the same preference, if voters could communicate among themselves then there would be no problem of aggregating information. Our theorems therefore identify common value environments for which communication is not necessary for information aggregation. In section 6.3, we show that our insights can also be applied to diverse value environments to identify necessary conditions for FIE to hold (without communication). It is no longer clear to what extent communication improves outcomes in the diverse value model.

Theorem 1 has opposite implications for discrete and continous state spaces. First, suppose that the state space is isomorphic to the simplex over signals, i.e., there are well defined

[1]We thank Timothy Feddersen for providing us this example.

[2]Formally, we impose that the set of events where the population would be indifferent has ex-ante zero probability.

preferences for all probability distributions over signals. In this case, Theorem 1 says that FIE holds if and only if $\mathcal{A}_1^\Delta$ and $\mathcal{A}_2^\Delta$ from a convex partition of the simplex. In other words, all the (indifferent) vectors around which the ranking changes must lie on a hyperplane. Therefore, if preferences were to be defined over a sufficiently rich space, information would be aggregated only in special environments.

At the other extreme, suppose that there are only two states. If $\mathcal{A}_1^\Delta$ and $\mathcal{A}_2^\Delta$ are two singletons, they can always be separated by a hyperplane. Therefore, in a two-state world which has been the focus of much of the literature, information is trivially aggregated. If there are three states and at least three signals, it is easy to check that we can separate any two probability vectors on the simplex from the third one by a hyperplane, as long as the three vectors are not collinear.

Theorem 2 presents a general condition on the information structure that is sufficient for an environment (with two or more alternatives) to allow FIE. The condition simply requires that the set of probability vectors arising in the different states be linearly independent. This result implies that FIE holds generically (Proposition 2) when state and signal spaces are of high cardinality, and also in discrete settings when there are at least as many signals as states (Corollary 1). Note also that those are conditions only on the information structure (mapping from states to the simplex) and not on preferences (mapping from states to rankings over alternatives).

When the state space is rich in the sense that it is isomorphic to the simplex (over finitely many signals) FIE holds for very special utility functions. Indeed, the linear independence condition (Theorem 2) is violated for rich state spaces. We provide a separate sufficient condition for FIE for this case. The environment in this situation is described by a partition $\mathcal{A}^\Delta$ of the simplex where each element $\mathcal{A}_j^\Delta$ is the set of conditional probability distributions for which a_j is the most preferred alternative. Convexity of $\mathcal{A}_j^\Delta$ for each alternative a_j, which is already demanding, is necessary but not sufficient for FIE when there are more than three alternatives. Theorem 3 says that an environment allows FIE if for every pair of alternatives a_i and a_j, the set of distributions for which a_i is preferred over a_j be separable from the set for which a_j is preferred over a_i.

Finally, we provide a representation theorem for environments that allow FIE. Proposition 1 says that (i) if the utility from each alternative is a linear function of the conditional probability vectors arising in the possible states, then the environment allows FIE; and (ii) conversely, for any environment that allows FIE, there exists another environment with the same top-ranked alternative in every state) which allows FIE and admits a utility that is linear in the probability vector. For rich state spaces, this means that the "marginal rate of substitution" between any two signals must be independent of the proportion of any other signal.

This result points to another source of aggregation failure: when the tradeoff between two sides of the issue under contention is affected by a third factor. Suppose a country is voting on whether to remain in an open trade zone or to exit (a recent important case would be the

British referendum in May 2016 on whether to continue in the Eurozone). A central tradeoff that drives voter preferences is that trade induces growth but leads to immigration as well, leading to loss of jobs for the local population. However, the tradeoff between growth and immigration depends on the extent to which immigrants contribute to the economy. Each voter receives a signal about one of these three factors, and the frequency of a signal depends on the strength of the factor. A similar issue arises in case of minimum wage legislation: the tradeoff of income and employment is affected by other factors like inflation. In all these cases, the information structure is too complex to guarantee that the correct outcome will prevail in all possible situations.

Our work speaks to a large body of literature on information aggregation in elections developed in the last three decades. The earlier statistical work (see Ladha (1992) and Berg (1993) among others) assumed that voters voted for the alternative that their private signals favored. Austen-Smith and Banks (1995) showed that such "sincere voting" is not necessarily rational when voters condition their decision on a tie, which is when their vote matters. The subsequent strand of game theoretic work (Feddersen and Pesendorfer (1997), Wit (1998), Myerson (1998), Duggan and Martinelli (2005), Meirowitz and Acharya (2013), McMurray (2016)) focussed on demonstrating that the *equilibrium* outcome of the voting game with large populations approximated the full information outcome while making specific assumptions on the informational and institutional environments (e.g., monotone likelihood ratio property, infinite signal spaces, supermajority rules, abstention, etc). Our approach has been to characterize the set of environments where FIE is achievable in (some) equilibrium, rather than making specific assumptions on the environment. Moreover, all these papers use a binary state space except for Feddersen and Pesendorfer (1997) (diverse preferences and ordered state space) and McMurray (2016) (common values and ordered state space). Thus, all except Feddersen and Pesendorfer (1997) can be seen as instances of our result that aggregation is generically possible when the state space is either discrete or ordered. We show that in order to better understand the problem of information aggregation in a non-trivial way, one must move away from the traditional two-state formulation. Finally, we show in section 5 that the characterization with majority rule remains unchanged if we allow scoring rules, supermajority rules, or abstention in order to connect with this line of work.

McLennan (1998) shows that if in a given environment there is a type-symmetric voting strategy profile that aggregates information, there is also a Nash equilibrium profile of strategies that aggregates information. We simply identify the environments where aggregation is feasible in this sense, and adapt McLennan's result to show that feasibility of FIE implies FIE in equilibrium (Theorem 4). Therefore, our paper brings forward McLennan's point that under common values, aggregation problems arise due to failure of feasibility and not due to perverse equilibrium inferences.

There is another strand of literature that identifies sources of aggregation failure in common values environments. This includes unanimity rules (Feddersen and Pesendorfer (1998)), alternative voter motivations (Razin (2003), Callander (2008)), cost of information acquisition

(Persico (2003), Martinelli (2005)), cost of voting (Krishna and Morgan (2012)), aggregate uncertainty (Feddersen and Pesendorfer (1997)), diverse preferences (Bhattacharya 2013) and so forth. Our work suggests that complexity of the information structure may itself be a barrier to information aggregation.

Mandler (2013) shows that aggregation can break down in a common values model if the same signal indicates opposite states in different situations. Bhattacharya (2013) presents a condition called Weak Preference Monotonicity which says that aggregation can fail if the same change in signal induces a randomly selected voter to vote for different alternatives for different beliefs over states. While these papers have an analogous message, our paper provides a stronger result in that we say that aggregation fails in all equilibria while they only identify particular bad equilibria.

There is a literature on informational efficiency on different scoring rules, in particular when there are three alternatives. While Goertz and Maniquet (2011) and Bouton and Castanhiera (2012) use diverse values, Ahn and Oliveros (forthcoming) have a common values model where they show that the approval rule performs weakly better than all other scoring rules. We show, conversely, that if a pure common values environment satisfies FIE for the approval rule, it does so for other scoring rules too.

There is also a parallel literature on information aggregation in common value auctions (Pesendorfer and Swinkels, 1997, Siga, 2016), and our Theorem 2 draws heavily from Siga (2016), which also provides a characterization of environments where information is aggregated in the auction context.

In our final section we show that Theorem 1 can be extended to include the case where voters have diverse preferences in addition to diverse information. However, in this case, we cannot apply McLennan's result to claim that whenever FIE is feasible, it is also achieved in equilibrium. Therefore, we only have general conditions for feasibility of FIE when preferences are diverse. Both Feddersen and Pesendorfer (1997) and Bhattacharya (2013) do satisfy the feasibility conditions.

The rest of the paper is organized as follows. Section 2 lays out the model with common voter preferences. Section 3 deals with the two-alternative case, where we provide a necessary and sufficient condition under which an environment allows full information aggregation. Section 4 provides sufficient conditions for FIE when there are multiple alternatives: We first provide a sufficient condition for general environments, and then provide one for rich state spaces where the previous condition fails. Section 5 shows that existence of a feasible strategy profile that fully aggregates information implies the existence of an equilibrium sequence of profiles that does the same. Section 6 discusses three extensions: the first one shows that our results do not change if we consider more general voting rules, the second one discusses the genericity of information aggregation, and the other one derives the conditions for existence of information-aggregating strategy profile in the case where voters may have preference heterogeneity. Section 7 concludes.

2. Model

Consider n players voting over k alternatives in $A = \{a_1, a_2, ..., a_k\}$.[3] We consider a plurality voting environment: each player chooses a vote in A, and the alternative with the largest number of votes wins the election. Ties, if any, are broken randomly. We consider alternative voting environments in section 5.

All voters have the same preference. The utility of a voter from an alternative depends on an unobservable state variable $\theta \in \Theta$, where Θ is a compact metric space. The generality of this formulation allows for both discrete and continuous state spaces, and help us to draw conclusions on the two distinct environments. The common utility of each voter is given by a bounded and continuous function $u : \Theta \times A \to \mathbb{R}$.

Let X be a compact metric space. We interpret X as the space of signals, and assume that, given a state θ, each voter privately draws an independent signal $x \in X$ according to a conditional probability distribution $P(\cdot|\theta) \in \Delta(X)$, where $\Delta(E)$ denotes the space of Borel probability measures over a topological space E. We assume existence of a Borel probability measure $\lambda \in \Delta(X)$ such that, for each $\theta \in \Theta$, $P(\cdot|\theta)$ is absolutely continuous with respect to λ. Let $f(\cdot|\theta)$ denote the corresponding density, and assume that $f : X \times \Theta \to \mathbb{R}$ is continuous.[4] We will abuse notation and use the same letter P to denote the prior probability on Θ.[5] Given our independence assumption, we informally refer to P as the information structure of the game.

We denote by $\mathcal{A}_j = \{\theta \in \Theta : u(\theta, a_j) \geq u(\theta, a_i) \text{ for all } i\}$ the set of states where alternative a_j is weakly preferred to all other alternatives. We assume: (i) $P(\mathcal{A}_j) > 0$ for all $j = 1, 2, ..., k$ (i.e., every alternative can be preferred ex-ante with positive probability); and (ii) $P(\mathcal{A}_i \cap \mathcal{A}_j) = 0$ for all $i \neq j$ (i.e., ex-ante indifference occurs with zero probability). The second assumption implies that both ex-ante and ex-post rankings are effectively strict.

A tuple $\{u, A, \Theta, X, P\}$ is defined as an environment. An environment in addition to an electorate size n defines a game. In a game, a strategy for a voter specifies a probability of voting for each alternative given each signal. We focus on symmetric strategies where voters with the same signal use the same strategy. A mixed strategy σ for a voter is a list of functions $\sigma_1, \sigma_2, ..., \sigma_k$ with $\sigma_j : X \to [0, 1]$ satisfying $\sum_j \sigma_j(x) = 1$ for all $x \in X$. In short, σ is a behavioral strategy mapping X to $\Delta(A)$, where $\sigma_j(x)$ is understood to be the probability of voting for the alternative a_j on obtaining signal x. When the context is clear we shall refer to σ as a profile of strategies, with the understanding that every player uses the same σ.

---

[3]We will use the words *voters* and *players* exchangeably depending on the context.

[4]Continuity of f, and even existence of such a density—that is, even the need for a dominating probability measure λ—can be relaxed in some of our results. In particular, most of our results hold true without either such condition, or with the weaker requirement that $\theta \mapsto P(\cdot|\theta)$ is strongly continuous. For ease of exposition, we will keep these assumptions throughout the paper, and note where we know weaker conditions would suffice.

[5]Of course, one could start from a probability measure $P \in \Delta(\Theta \times X)$ and derive conditionals and marginals. Since we will have no use for such an underlying probability, we work directly with these three concepts as primitives.

Given σ, the expected vote share of alternative a_j at state θ is given by

$$z_j^\sigma(\theta) = \int_X \sigma_j(x) P(dx|\theta).$$

By the SLLN, when every voter uses the strategy σ then, for every θ, the realized proportion of votes for alternative a_j converges $P(\cdot|\theta)$-a.e. to $z_j^\sigma(\theta)$ as $n \to \infty$. Since our focus is on large electorates, we call $z_j^\sigma(\cdot)$ the vote share function for alternative a_j. The vote share function for each alternative has two properties: first, it is continuous in θ; and second, it is linear in $P(\cdot|\cdot)$, i.e., in the vectors lying on the simplex $\Delta(X)$.

Next, we define the standard for information aggregation for a given strategy profile.

2.1. **Full Information Equivalence.** In a large electorate, the signal profile almost surely reveals the state. Thus, if the signal profile were publicly observed, the most preferred alternative would almost surely be elected. We say that information is aggregated by a strategy profile if, under private information, the most preferred alternative is guaranteed to win with an arbitrarily high probability. We formalize this idea now.

Given any strategy profile σ and electorate $\{1, ..., n\}$, let z_n^σ denote the *realized* vector of proportion of votes for alternatives $a_1, ..., a_k$. Observe that θ and σ induce a probability distribution p_θ^σ over z_n^σ, since the signal profile is drawn according to $P(\cdot|\theta)$ and given the realized signal profile, the profile of votes is drawn according to σ. Formally, given a strategy profile σ and a profile $x_1, ..., x_n$ of signals, the probability of a proportion $y = (y_1, ..., y_k)/k$, with $y_j \in \{1, ..., n\}$ for $j = 1, ..., k$ and $\sum_{j=1}^k y_j = k$, is given by

$$p_n^\sigma(y|x_1, ..., x_n) = \sum_{\mathcal{B}(y)} \prod_{\ell=1}^k \prod_{i_\ell \in B_\ell} \sigma_\ell(x_{i_\ell})$$

where $\mathcal{B}(y) \equiv \{(B_1, ..., B_k) : (B_1, ..., B_k) \text{ is a partition of } \{1, ..., n\} \text{ with } |B_j| = y_j, j = 1, ..., k\}$, and for any set Z, $|Z|$ is the number of elements in Z. Then the probability of y given σ and θ is

$$p_n^\sigma(y|\theta) = \int p_n^\sigma(y|x_1, ..., x_n) \otimes_{i=1}^n P(dx_i|\theta).$$

Let L_n^j denote the set of vectors of proportions where the jth coordinate is *not* the unique highest. A wrong outcome is obtained if, in a state where a_j is the most preferred alternative, it fails to garner the unique maximum number of votes. Thus, the ex-ante probability of obtaining a "wrong" outcome is

$$W_n^\sigma = \sum_j \int_{\mathcal{A}_j} p_n^\sigma(L_n^j|\theta) P(d\theta).$$

We say that informaton is fully aggregated if $W_n^\sigma \to 0$ as $n \to \infty$. More formally, we say that in an environment $\{u, A, \Theta, X, P\}$, the strategy σ achieves Full Information Equivalence (FIE) if the ex-ante likelihood of error induced by σ converges to 0 as the number of voters increases unboundedly.

Alternatively, for a given σ, let us define the set of states where alternative a_j is elected by

$$\mathcal{A}_j^\sigma = \{\theta : z_j^\sigma(\theta) > z_i^\sigma(\theta), \text{ for } i \neq j\}.$$

We say that σ achieves Full Information Equivalence (FIE) if the set of states where the preferred alternative fails to win almost surely is of m-measure zero. In other words, FIE obtains if for all j,

$$P(\mathcal{A}_j \setminus \mathcal{A}_j^\sigma) = 0.$$[6]

Observe that we restricted ourselves to symmetric strategies, that is, to the case that each voter uses the same common strategy σ. One can extend the definition of FIE and allow for sequences of strategies that are not necessarily composed of the same common strategy. Not much is gained by that because of the following: if the sequence $(\sigma^1, ..., \sigma^n, ...)$ achieves FIE, then the common strategy σ defined by

$$\sigma(x) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sigma^m(x) \text{ for } P(\,\cdot\,|\theta)\text{-a.e. } x$$

also achieves FIE, provided that such limit exists.

There is no guarantee that a strategy ensuring full information equivalence will exist for every environment. In fact, if there exists some strategy that aggregates information in a given environment, we say that the environment *allows* FIE.

The next section discusses properties of the environments that allow FIE. Notice that an environment allowing FIE is necessary but not sufficient for information aggregation in equilibrium.

3. Feasibility of Information Aggregation: Two Alternatives

In this section, we discuss the special case of two alternatives, i.e., $k = 2$. Starting with the case of binary alternatives has several advantages. First, this case is of great interest in itself. Aside from empirical relevance, almost the entire literature on CJT concentrates on the binary alternatives case. As a basis of comparison with the rest of the literature, it is useful to study this case separately. Second, with two alternatives, we obtain a sharp characterization of the environments that allow FIE by deriving a condition that is both necessary and sufficient. This condition is very useful for understanding the issues that drive FIE. Finally, the binary case serves as a starting point for understanding the more general case of multiple alternatives.

To demonstrate the condition that determines whether an environment allows FIE or not, we start with two examples. In the first example we show how FIE may fail when we depart

[6] To verify that the two definitions are equivalent, say that $P(\mathcal{A}_j \setminus \mathcal{A}_j^\sigma) = 0$ for all j. Then $W_n^\sigma = \sum_j \int_{\mathcal{A}_j^\sigma \cap \mathcal{A}_j} p_\theta^\sigma(L_n^j) P(d\theta)$. For θ in the domain of integration, we have $z_j^\sigma(\theta) \geq z_i^\sigma(\theta)$ for every i, and we know that the realized proportion $z_j^n(\theta)$ converges to $z_j^\sigma(\theta)$, so $p_\theta^\sigma(L_n^j) \to 0$ for every j, establishing that $W_n^\sigma \to 0$ as $n \to \infty$. Conversely, if $P(\mathcal{A}_j \setminus \mathcal{A}_j^\sigma) > 0$ for some j, then there is a set of positive P-measure $E \subset \mathcal{A}_j$ and an alternative i such that $z_i^\sigma(\theta) > z_j^\sigma(\theta)$ for all $\theta \in E$. Because the realized vector of proportions converges to $z^\sigma(\theta)$, $p_\theta^\sigma(L_n^j)$ does not vanish for $\theta \in E$; but this means that W_n^σ does not converge to 0.

from the most standard framework in a Condorcet Jury environment. Most of the literature impose that there is a unique state of the world, and a unique distribution of signals where an alternative is best. For instance, when there are two alternatives it is standard to assume that there are two distinct distributions of signals, one where an alternative is best, and another distribution for the other alternative. We show that FIE may not be reached when an alternative may be best in multiple states of the world. The second example develops similar ideas in a richer state space.

Let $A = \{a_1, a_2\}$, and $\sigma = (\sigma_1(\cdot), \sigma_2(\cdot))$, where $\sigma_j(x)$ is the probability of voting for alternative a_j on obtaining signal $x \in X$. Given the strategy, we write the difference in expected vote share between a_1 and a_2 as

$$\begin{aligned} z_{12}^{\sigma}(\theta) &\equiv z_1^{\sigma}(\theta) - z_2^{\sigma}(\theta) \\ &= \int_x [\sigma_1(x) - \sigma_2(x)]\, P(dx|\theta) \equiv \int_x \sigma_{12}(x) P(dx|\theta) \end{aligned} \tag{1}$$

where $\sigma_{12}(x) = \sigma_1(x) - \sigma_2(x)$ for all $x \in X$.

In order for FIE to obtain, the function $z_{12}^{\sigma}(\theta)$ should be positive in states belonging to $\mathcal{A}_1$ (where a_1 is preferred) and negative in states belonging to $\mathcal{A}_2$ (where a_2 is preferred).

Example 1. *Suppose $A = \{a_1, a_2\}$, $\Theta = \{L, M, R\}$, and $P(\theta) = \frac{1}{3}$ for all $\theta \in \Theta$. Assume a_1 is preferred in L and R while a_2 is preferred in M. Also, $X = \{x, y\}$, and for some $p > \frac{1}{2}$,*

$$P(x|L) = p, \ \Pr(x|M) = \frac{1}{2}, \ \textit{and } P(x|R) = 1 - p$$

This environment does not allow FIE.

This proof and all the reminder of the proofs are relegated to the appendix. The problem with information aggregation in this example is the following: In order for a_1 to win in state L where x is the more frequent signal, voters with signal x should vote for a_1 with sufficiently high probability. Similarly, in order for a_1 to win in state R where y is the more frequent signal, voters with signal y should vote for a_1 with sufficiently high probability. But then, for both the signals, voters vote for a_1 more frequently than for a_2. As a consequence, in state M where both these signals occur with moderate probability, the vote share for a_1 is already high and a_2 cannot win.

Example 2. *Let $A = \{a_1, a_2\}$, $\Theta = [0, 1]$ with a uniform prior probability, $X = \{x, y\}$, and $\Pr(x|\theta) = \theta$. Consider two different preference environments. In the first case, suppose all voters prefer a_1 if $\theta > t$ and a_2 for $\theta < t$, for some $t \in (0, 1)$. In the second case, for some $0 < t_1 < t_2 < 1$, a_1 is preferred whenever $\theta \in (t_1, t_2)$ and a_2 is preferred when $\theta < t_1$ or $\theta > t_2$. The first environment allows FIE but the second environment does not.*

Example 2 has the following real world implication. Suppose the voters have only one of two kinds of information. If signals are about relative quality of the candidates, one signal is interpreted as positive information about a_1 and the other about a_2. However, if signals are about whether a candidate leans to the left or right, and voter preferences are depend on whether the candidate is moderate or extreme, signals cannot be classified as each favoring one candidate. Aggregation fails in this second case as well.

The main idea underlying the characterization theorem is contained in example 2. In this example, FIE depends on the convexity of the set of states for which a given alternative is preferred: in the first environment both the sets $\mathcal{A}_1$ and $\mathcal{A}_2$ are convex, while in the second environment the set $\mathcal{A}_2$ is non-convex. The example, however, is special because the state space is isopmorphic to the simplex of signal probabilities. Therefore, convexity of the sets $\mathcal{A}_1$ and $\mathcal{A}_2$ are incidental to the example: in general, the condition for an environment allowing FIE is a convexity condition on the set of the probability distributions where a particular alternative is preferred. Theorem 1 formalizes the idea.

3.1. **Characterization for two alternatives.** Before stating the main result of this section we introduce some necessary notation. We say that a state $\theta \in \Theta$ is *pivotal* if, for each $\varepsilon > 0$, $P(\mathcal{A}_1 \cap B_\varepsilon(\theta))$ and $P(\mathcal{A}_2 \cap B_\varepsilon(\theta))$ are positive, where $B_\varepsilon(\theta)$ is an ε open ball around θ. Let $M^{piv} \subset \Theta$ denote the set of pivotal states. By continuity of the utility function, at all pivotal states, the voters must be indifferent between a_1 and a_2. Given our assumption that indifference occurs only with zero probability, pivotal states occur only when the state space is continuous.

A hyperplane on $\Delta(X)$ is denoted by $H = \{\mu \in \Delta(X) : \int_X h(x)\mu(x) = 0\}$, for a given measurable function $h : X \to \mathbb{R}$. Given a hyperplane H, we use $H^+ = \{\mu \in \Delta(X) : \int_X h(x)\mu(x) > 0\}$ and $H^- = \{\mu \in \Delta(X) : \int_X h(x)\mu(x) < 0\}$ to denote the two associated half-spaces. We shall denote the vector h as the normal to the hyperplane H.

Theorem 1. *Suppose $A = \{a_1, a_2\}$. An environment (u, A, Θ, X, P) allows FIE if and only if there exists a hyperplane H in $\Delta(X)$ such that $P(\cdot|\theta) \in H^+$ for $\theta \in \mathcal{A}_1$, $P(\cdot|\theta) \in H^-$ for $\theta \in \mathcal{A}_2$, and, if $M^{piv} \neq \emptyset$, $P(\cdot|\theta) \in H$ for $\theta \in M^{piv}$.*

Here is important to note that the vote share function for each alternative has two useful properties: first, it is continuous in θ; and second, it is linear in $P(\cdot\,|\,\cdot)$, i.e., in the vectors lying on the simplex $\Delta(X)$. The intuition for necessity follows example 2. For FIE, it must be the case that, that $z^\sigma_{12}(\theta)$ has to be positive in the $\mathcal{A}_1$-states and negative in the $\mathcal{A}_2$-states. Since $z^\sigma_{12}(\theta)$ is a linear functional of the conditional probability vectors, the respective images of the set of $\mathcal{A}_1$-states and the $\mathcal{A}_2$ -states on the simplex must be convex (or, contained in disjoint convex hulls). In particular, for any pivotal state, we must have $z^\sigma_{12}(\theta) = 0$ by continuity. Another way to see this is by noting that σ_{12} on equation (1) can be interpreted as the norm of a separating hyperplane. For sufficiency, suppose there is a hyperplane with normal h on the simplex that separates the distributions arising in $\mathcal{A}_1$-states from those in the $\mathcal{A}_2$-states. Any strategy profile with σ_{12} proportional to the normal h will deliver FIE.

It must be noted that while we cannot uniquely identify the strategies that aggregate information, all strategy profiles that aggregate information are characterized by a specific property: in pivotal states, they lead to equal share of votes. In section 5, we show that in environments that allow FIE, there exists also an equilibrium strategy profile with the same properties in the limit. McMurray (2014) finds a result with a similar flavor and interprets it as the endogenous emergence of a single dimension of political conflict in a multidimensional world.

3.2. **Discussion of Theorem 1.** A concise way to describe the theorem is as follows. Denote by $\mathcal{A}_j^\Delta$ the image of $\mathcal{A}_j$ on $\Delta(X)$, i.e., $\mathcal{A}_j^\Delta = \{P(\cdot|\theta) \in \Delta(X) : \theta \in \mathcal{A}_j\}$. Theorem 1 says that the environment allows FIE if and only if $\mathcal{A}_1^\Delta$ and $\mathcal{A}_2^\Delta$ are separated by a hyperplane. Figure 1a illustrates an environment with discrete states that allows FIE. The blue dots are distributions where, say, a_1 is preferred, while the red dots a_2 is preferred.

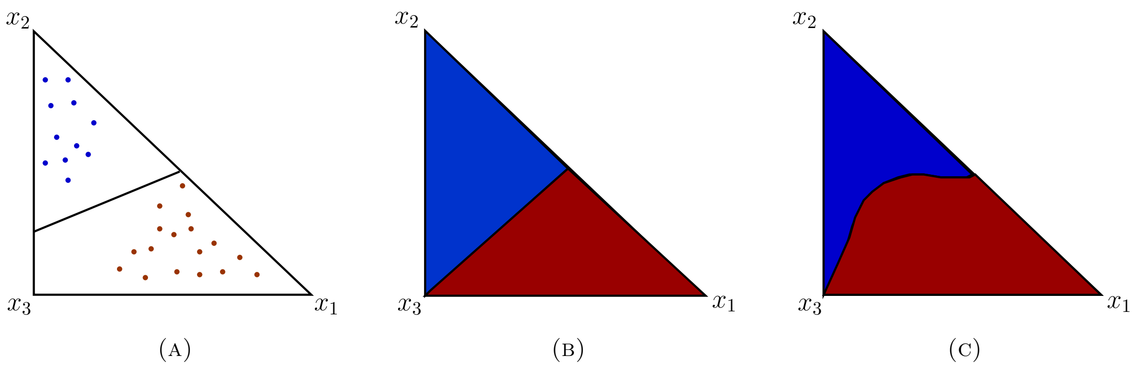

FIGURE 1. Illustrating Theorem 1.

Theorem 1 offers a clear interpretation of signals in environments with two alternatives that allow FIE. In any such environment, the set of signals X can be divided into two classes X_1 and X_2, in the sense that those signals in X_1 (resp. X_2) have their corresponding vertices in the upper halfspace of a hyperplane on the simplex where a_1 (resp. a_2) is preferred. Signals in X_1 favor a_1 and those in X_2 favor a_2 in the following sense: for any signal in X_1, the strategy that achieves FIE attaches higher probability to a_1 than to a_2 ($\sigma_{12}(x) > 0$ if $x \in X_1$) and for any signal in X_2,the strategy attaches higher probability to a_2. In figure 1a, for example, the line dividing the blue and red states puts two vertices on the red side (i.e., the vertices for signals x_1 and x_3). With a strategy profile that generates the line in the graph, the higher the probability of signals x_1 and x_3, the more votes a_2 receive. However, we can alternatively draw a line that puts only the x_1 vertex on the red side, and thus, x_3 favors a_1. Therefore, in case of discrete state spaces, signals endogenously favor alternatives. In the case of a continuous state space giving rise to a dense set of conditional probability distributions on the simplex, the classification of signals may be unique. Figure 1b shows an example where all distributions over signals may arise, say, with uniform probability. In this case there is a unique separation of signals congruent with FIE.

Theorem 1 can be seen as a positive or a negative result for the genericity of information aggregation depending on the class of information structures that one considers. To see the idea, first consider a continuous state space so that there are pivotal states. The theorem says that the images of all pivotal states must lie on a hyperplane on the simplex. Thus, FIE imposes strong restrictions on the utility functions when the state space is continuous. A particular case of importance is when preferences are defined over all probability distributions over signals, where the hyperplane condition is satisfied only in special environments. One must remember that in an environment that does not allow FIE, information cannot be aggregated even if voters can co-ordinate on a strategy profile.

Example 3 simply extends example 2 from a unidimensional to a multidimensional simplex over signals.

Example 3. *Let* $X = \{x_1, x_2, x_3\}$*, and consider a uniform prior distribution on* $\Delta(X)$*, where for all* $\mu \in \Delta(X)$*,* μ_i *is the probability of signal* x_i*. Suppose, as in example 2, that preferences are defined directly on* $\Delta(X)$*, and let the function* $v_{12} : \Delta(X) \to \mathbb{R}$ *denote the utility difference between* a_1 *and* a_2*. Suppose that the signal* x_1 *is optimistic about* a_1 *(albeit pessimistic about* a_2*), and vice versa for the signal* x_2*. It is reasonable given these interpretations that* $v_{12} \cdot \mu > 0$*, for* $\mu \in \Delta$*, when* $\mu_1 = 1$ *(i.e., in a state where the signal* x_1 *is observed with probability 1). Likewise,* $v_{12} \cdot \mu < 0$ *for* $\mu \in \Delta$*, when* $\mu_2 = 1$*. It is also reasonable that along the path of indifference, more* x_1 *signals would compensate for more* x_2*. Theorem 1 implies that the only type of indifference between the two alternatives on* $\Delta(X)$ *that allow FIE is a straight line. Figure 1b illustrates a case that allows FIE, while a more generic indifference as in figure 1c will not aggregate information.*

To fix the idea of example 3, consider an electorate voting on a referendum, and signals provide information about the proposal. One can think of x_1 and x_2 being information on two sides of a tradeoff that the proposal entails. For example, in case of a vote over whether to remain in a common politico-economic union or not (e.g., the "Brexit" vote in May 2016). Here, x_1 is a signal that says that that staying will be good for growth and x_2 is a signal that says there will be loss of local jobs due to migration. The stronger the likely growth effect is, the more x_1 signals are received, and the larger the likely job loss is, the more x_2 signals are received. Now, think of x_3 being a signal on a third factor, say, the extent to which migrants make a net contribution to the local economy. Theorem 1 says that the proportion of $x = 0$ signals should not affect the rate at which μ_1 (the proportion of signals about growth) is traded off against μ_2 (the proportion of signals about job loss). This seems to be a strong restriction on preferences.

The next example is an application of Theorem 1 to a very standard case of spatial model of political competition between two alternatives.

Example 4. *We continue the metaphor of a policy proposal (alternative* a_1*) being voted on against a status quo (alternative* a_2*). There is a policy space* $Y = [0, 1]^2$*, in which both alternatives are located. Voter utility for policy* y *is given by* $u(|y - y^*|)$*,* $u' < 0$*. Thus,* $y^* \in Y$ *is the voter ideal policy and voters prefer policies closer to* y^* *than those further from it. The status quo is known to be located at* $y_Q \neq y^*$ *on the policy space. On the other hand, there is uncertainty about the location of the proposed policy: we denote the location of the proposed alternative on the policy space by* $\theta = (\theta_1, \theta_2) \in [0, 1]^2$*. In this setting, the voters prefer* a_1

(resp. a_2) in a given state θ if $|\theta - y^*|$ is less (greater) than $|y_Q - y^*|$. The pivotal states are given by

$$\mathcal{I} = \{\theta : |\theta - y^*| = |y_Q\text{-}y^*|\} \tag{2}$$

which is the circumference of a circle (or part thereof). This is entirely a property of the utility function. Suppose the prior probability of the alternative is uniform on the policy space, the signal $x = (x_1, x_2)$ is two dimensional, and $x_i \in \{0, 1\}$, with $P(x_i = 1|\theta_i) = \theta_i$. Thus, x_1 provides information on θ_1 and x_2 on θ_2 independently of each other. Alternatively, the state θ_i can simply be thought of as the proportion of 1-signals in dimension i. There is no strategy profile for which FIE can be obtained in this setting.

Both examples 3 and 4 feature rich (continuous) state spaces but finite number of signals. In these examples, voters have very limited private information and FIE is not satisfied. Therefore, if voters have limited private information but their preferences have a rich variation across different circumstances (states), then even with two alternatives it can be impossible to guarantee information aggregation despite common preferences.

The following two examples presents the polar opposite case: the state space is discrete (and there are at least as many signals as states). In this case, we shall see that irrespective of the particular utility function, FIE obtains *except* for special circumstances, even when there are more than two alternatives in question. We shall develop the result more generally in the next section (Corollary 1). The examples in this section simply illustrate the idea. As the examples make it clear, the crucial difference between continuous-state and discrete-state environments is that in the former, FIE requires a strategy to produces equal vote shares for the two alternatives at all pivotal states but there is no such strict requirement in the latter.

In the two examples below, there are r states $\{\theta_1, \theta_2, ..., \theta_r\}$ occuring with positive probability and s signals. The model assumptions imply that in each state, the ranking over alternatives is strict. Each state θ_t leads to a distinct probability distribution $P(\cdot|\theta_t)$ denoted as P_t.

Example 5. *FIE obtains whenever $r = 2$ and $P_1 \neq P_2$.*

Example 6. *FIE obtains if $r = s = 3$ and the conditionals arising in the three states $\{P_1,\ P_2,\ P_3\}$ are linearly independent.*

We skip the proof as these are special cases of Corollary 1 introduced later. However, for these cases, the statements can simply be verified by inspecting the figures. Figure 2a illustrates the case of $r = 2$: it is always possible to pass a hyperplane separating P_1 and P_2 irrespective of their location on the simplex. Notice that this example is of independent interest given that there is a large literature that looks specifically at this case.

Similarly, Figure 2b illustrates the case of $r = 3,\ s = 3$ when the conditional vectors satisfy linear independence. It is easy to see from the figure that one can always separate any two vectors from the third by a suitable hyperplane. Figure 2c demonstrates the necessity of linear independence. In this case, if a_1 is preferred in θ_1 and θ_3 while a_2 is preferred in θ_2.

It is easy to see from the figure that one cannot find a hyperplane that separates $\{P_1, P_3\}$ from P_2.

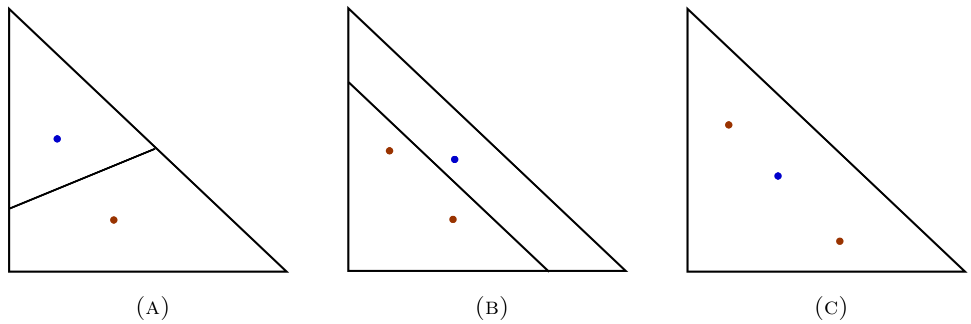

FIGURE 2. Discrete state space.

4. Multiple alternatives: Feasibility

In this section, introduce sufficient conditions for FIE when there are more than two alternatives. The two-alternative case is special in that it is straightforward to characterize the conditions for FIE. In general however, such conditions are hard to obtain. We provide two sets of results. The first set of results provides a simple sufficient condition for FIE in general informational environments. However, the first set of conditions do not apply to the case of rich state spaces (in the sense that the state space is isomorphic to the simplex over signals). We develop a separate, set of conditions for rich state spaces.

Recall that $\mathcal{A}_j$ is the set of states where alternative a_j is (weakly) the most preferred, and $\mathcal{A}_j^\Delta$ is its image on the simplex over signals. We denote the partition of the state space $\{\mathcal{A}_j\}_{j=1}^k$ by $\mathcal{A}$ and, correspondingly, $\{\mathcal{A}_j^\Delta\}_{j=1}^k$ by $\mathcal{A}^\Delta$.

While the rest of the paper analyzes sufficient conditions for different environments, the next remark identifies the necessary condition

Remark 1. *A necessary condition for an environment to allow FIE is that, for each for each* $i \neq j$, $\mathcal{A}_i^\Delta$ *and* $\mathcal{A}_j^\Delta$ *are separable by a hyperplane on the simplex.*

We skip the proof as it is immediate, and follows the same logic as Theorem 1.

4.1. Sufficiency condition for General Information Environments.

Our next task is to find a sufficient condition for FIE in environments with multiple alternatives. At this point, it would be important to recall that the class of admissible information structures $\mathcal{F}$ comprises of joint distributions P over $\Theta \times X$ that are absolutely continuous with continuous densities. Notice that when the state and signal spaces are discrete, these assumptions are trivially satisfied.

Example 6 shows that for three states, FIE is always satisfied if the set of conditionals is linearly independent. We now show that this is an instance of a more general result: as long as the information structure P satisfied a form of independence, the environment allows FIE. Formally we use the notion of independence introduced by McAfee and Reny (1992):

Definition 1 (Independence)**.** *We say that $P \in \mathcal{F}$ satisfies independence when the following condition holds true: if*

$$\int P(\,\cdot\,|\theta)\nu(d\theta) = P(\,\cdot\,|\theta)$$

for some probability measure $\nu \in \Delta(\Theta)$ then it must be that $\nu = \delta_\theta$, where δ_θ is the point-mass concentrated at θ.

When the state space is discrete, i.e., $\Theta = \{\theta_1, \theta_2, ..., \theta_r\}$ and the conditional $P(\cdot|\theta_t)$ in the generic state θ_t is denoted by P_t, independence boils down linear independence: There does not exist r scalars $\{\nu_1, \nu_2, ..., \nu_r\}$ with at least one not equal to zero such that

$$\sum_{t=1}^{r} \nu_t P_t = 0$$

The next theorem says that independence is sufficient for an environment to allow FIE. Notice that this is a property of the information structure: as long as the joint distribution P on $\Theta \times X$ satisfies independence, FIE obtains irrespective of the preference mapping from states to rankings over alternatives. We make heavy use of the techniques developed in the auction context by Siga (2016) in order to prove the theorem.

Theorem 2. *If the information structure $P \in \mathcal{F}$ satisfies independence, then the environment allows FIE.*

We develop the proof in two steps. First, we present a property which we call PS. An environment is said to have this property if we can find a set of *parallel* hyperplanes that separate each $\mathcal{A}_i^\Delta$ and $\mathcal{A}_j^\Delta$ (for $i \neq j$). Lemma 1 shows that this if an environment satisfies this property, we can construct a strategy that satisfies FIE. Lemma 2 shows that affine independence implies Property PS.

Definition 2. *We say that **Property PS** holds if there exists $h : X \to [-1, 1]$ and $0 < c_0 < c_1 < c_2 < \cdots < c_k$ such that for all $j = 1, ..., k$, $\mathcal{A}_j \subseteq \{\theta : \int h(x)P(dx|\theta) \in (c_j^{-1}, c_{j-1}^{-1})\}$.*

Lemma 1. *If property **PS** holds, then there exists a strategy that achieves FIE.*

The intuition for the proof is the following. The property PS means that each $\mathcal{A}_l^\Delta$ and $\mathcal{A}_j^\Delta$ are separated by a hyperplane H_{ij} on the simplex, and all the hyperplanes have a common normal, given by the vector $h(x)$. We show that we can define a strategy function σ such that $\sigma_{ij}(x) = h(x) - \frac{1}{c_i}$ for all x, and all $i \neq j$. This strategy function achieves the required separation and delivers FIE. This Lemma is based on Theorem 2 in Siga (2016).

Lemma 2. *If $P \in \mathcal{F}$ satisfies independence then property **PS** holds.*

In the proof, we show that the information structure P satisfies independence, we can always slice the simplex by parallel hyperplanes such that each slice contains the set $\mathcal{A}_j^\Delta$

corresponding to a given alternative a_j. Lemma 1 and Lemma 2 together prove the statement of Theorem 2.[7]

Theorem 2 allows a simple corollary for the case when both the state and signal space are finite.

Corollary 1. *Suppose there are r states $\{\theta_1, \ ..., \theta_r\}$, k alternatives and s signals with $s \geq r \geq k$. The environment allows FIE if the conditional vectors are independent.*

The above corollary generalizes examples 5 and 6. Of course, when $s \geq r$ generically we will have independence (we discuss this in more details in section 6). The general lesson with finite state and signal spaces is that FIE obtains except for very special cases as long as the number of signals is larger than the number of states, i.e., the signal space is richer than the state space. Corollary 1 is important because a large number of papers in the literature on the topic have concentrated on this set-up.

If on the other hand, the number of states is larger than the number of signals, then FIE is no longer a generic property. To see this, consider $r > s$ and $k = 2$. Then, the condition for FIE is given by Theorem 1. Now, suppose that the conditional distribution in each state is a randomly chosen vector on the s-simplex. As the number of states r goes up keeping s fixed, the likelihood of the the condition in the Theorem being violated increases. In fact, one can make the likelihood of FIE obtaining arbitrarily small by increasing the number of sates while holding the number of signals fixed (see Siga, 2016, theorem 3 for a general analysis). As an extreme case, example 3 and example 4 demonstrate that if there are an infinite number of states and a finite number of signals then, FIE fails generically.

4.2. **Rich State Spaces.** There is one class of important information structures that violate the sufficient conditions in Theorem 2. This is the case where the state space is isomorphic to the signal space. We have seen that in this case, Theorem 1 puts strong restrictions on the set of preferences that allow FIE (ref examples 3 and 4). This case is also of independent theoretical interest as it does not restrict the set of signal distributions over which voter preferences are defined. In this section we develop sufficient conditions for FIE when the state space is isomorphic to the simplex over signals, i.e., when preferences are defined over the entire simplex. Formally, let $\mathring{\Delta}(X) \subset \Delta(X)$ denote the set of probability measures in $\Delta(X)$ with full support and let supp$P$ denote the support of P on Θ. We then make the following assumption, used for the remainder of this section.

> **Assumption M:** The mapping M : supp$P \to \mathring{\Delta}(X)$, given by the restriction of $\theta \mapsto P(\,\cdot\,|\theta)$ to supp$P$, is a bijection.

---

[7] Here we show that the condition for sufficiency cannot be relaxed further to convex independence from (linear) independence. Suppose $r = s = 4$ and $k = 2$. The conditionals $P_1 = (0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $P_2 = (\frac{1}{2}, 0, 0, \frac{1}{2})$, $P_3 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$, and $P_4 = (0, 0, 0, 1)$ satisfy convex independence; assume that $\mathcal{A}_1 = \{\theta_1, \theta_2\}$ and $\mathcal{A}_2 = \{\theta_3, \theta_4\}$. It is clear that the sets $\{P_1, P_2\}$ and $\{P_3, P_4\}$ cannot be separated by a hyperplane. In fact, the vector $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5})$ is equal to $\frac{3}{5}P_1 + \frac{2}{5}P_2$ and also to $\frac{3}{5}P_3 + \frac{2}{5}P_4$, so it lies in the convex hull of each of these two sets; thus we cannot separate these two convex sets.

We can now write the utility function defined over the states as $\tilde{u} : \Delta(X) \times A \to \mathbb{R}$ such that, $\tilde{u}(P(\,\cdot\,|\theta), a) := u(\theta, a)$, for all $\theta \in \Theta$, and all $a \in A$. Given the isomorphism, we will refer to the prior over $\Delta(X)$ as P with some abuse of notation. Similarly, we shall write $z_a^\sigma(\mu)$ to mean $z_a^\sigma(\theta)$ where $P(\,\cdot\,|\theta) = \mu$. We also assume that the signal space is finite $X = \{x_1, x_2, ..., x_s\}$ for this section only.[8] Notice also that when the state space is rich, not only does $\mathcal{A}$ describe a partition of the state space, its image $\mathcal{A}^\Delta$ also describes a partition of the simplex.

An adaptation of Remark 1 tells us that the necessary condition for FIE for rich state spaces is that $\mathcal{A}_j^\Delta$ is convex for all j. As argued earlier, this condition seems rather strong and will be satisfied only in special cases. This result stands in stark contrast to the genericity of FIE for finite state spaces (see Corollary 1).

The next example shows that a convex partitional structure is not sufficient for FIE.

Example 7. *Suppose there are three signals and the simplex over the signals is represented by the right angled tiangle ABC, as shown in Figure 3 below. There is a smaller right angled triangle DEF inside ABC, with side EF parallel to BC. The line EF intersects AC at G. The most favored alternatives for different probability vectors in the triangle are as follows: a_1 for the trapezium $ADEB$, a_2 for the trapezium $ADFG$, a_3 for the trapezium $GEBC$, and a_4 for the triangle DEF. While each $\mathcal{A}_j^\Delta$ is convex, FIE is not achievable in this environment. To see why, suppose strategy σ achieves FIE. Now, it must be the case that along all points on AD (and the entire line along AD on the simplex), $z_1^\sigma(\mu) = z_2^\sigma(\mu)$. Similarly, for all points on the line along BE, $z_3^\sigma(\mu) = z_1^\sigma(\mu)$. By linearity of vote shares in μ, if BE and AD intersect at H, then at $\mu = H$, $z_1^\sigma(\mu) = z_2^\sigma(\mu) = z_3^\sigma(\mu)$ Similarly, at F which is the intersection of DF and EF, we must have $z_4^\sigma(\mu) = z_2^\sigma(\mu) = z_3^\sigma(\mu)$. By linearity, $z_{23}^\sigma(\cdot) = 0$ must trace a line on the simplex, but we already know two points on this line: H and F. Therefore, $z_{23}^\sigma(\cdot) = 0$ must be represented by the line along FH. However, for FIE we need $z_{23}^\sigma(\cdot) = 0$ to coincide with the line through GF, which is impossible.*

There are actually two reasons why the convex partitional structure may not be sufficient for FIE. Suppose, for each pair (i, j), the sets $\mathcal{A}_i^\Delta$ and $\mathcal{A}_j^\Delta$ are separated by the hyperplane with norm h_{ij}. By the logic of pivotal states for Theorem 1, we need to find a strategy σ which induces $z_{ij}^\sigma(\mu) = 0$ along the hyperplane that separates $\mathcal{A}_i^\Delta$ and $\mathcal{A}_j^\Delta$. In other words, σ_{ij} must be proportional to h_{ij} for all pairs (i, j). Now, when there are k alternatives, there are potentially ${}_kC_2$ such hyperplanes h_{ij}. On the other hand, a choice of a strategy function $\sigma = \{\sigma_1, \sigma_2, ...\sigma_k\}$ is essentially a choice of of k vectors (each with s elements). Therefore, we have k unknowns with which to satisfy, potentially, ${}_kC_2$ restrictions, and that cannot be done unless we impose extra conditions. Notice that with $k = 2$, we had one hyperplane and we could choose two vectors σ_1 and σ_2 to "match" a separating hyperplane. Therefore, with $k = 2$, convexity of $\mathcal{A}_j^\Delta$ was also sufficient for FIE. However, this problem is no longer trivial if $k > 3$ when there are potentially more hyperplanes than alternatives. The second reason that complicates the problem is that for any strategy σ, the hyperplanes described by

[8] Hence, in this case, the notation $\mathring{\Delta}(X)$ does mean the interior of the simplex $\Delta(X)$ as a subset of $\mathbb{R}^{s-1}$.

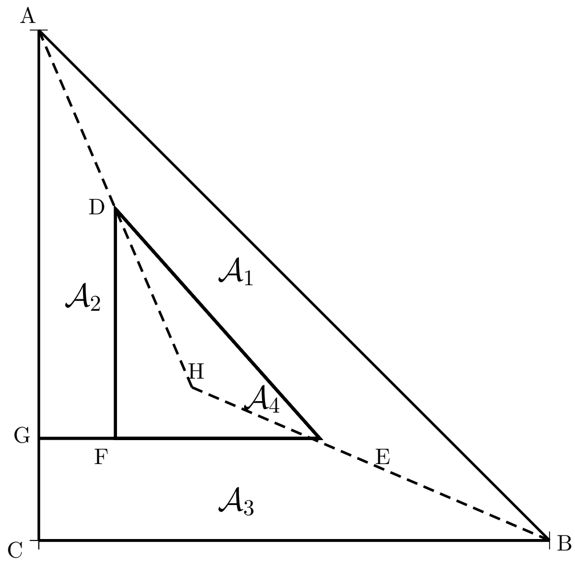

FIGURE 3. Convex partition.

$z_{ij}^{\sigma}(\mu) = 0$ for different pairs (i, j) are not necessarily linearly independent. This is already evident from example 7.

In order to develop further conditions that guarantee FIE, we extend the domain of preferences from the simplex, which is $\Delta = \{\mu \in \mathbb{R}^s : \mu \cdot \mathbf{e} = 1,\ \mu \geq 0\}$ to the space $\Sigma = \{\mu \in \mathbb{R}^s : \mu \cdot \mathbf{e} = 1\}$, which is the space of s-vectors that sum to 1.

Assumption E: There exists a utility function $\hat{u} : \Sigma \times A \to \mathbb{R}$ such that $\tilde{u} : \Delta(X) \times A \to \mathbb{R}$ is the restriction of $\hat{u}$ to $\Delta(X) \times A$.

The next assumption is the key property that guarantees sufficiency.

Assumption H: For all $a_i, a_j \in A$, there exists a hyperplane H_{ij} defined by the normal vector $h_{ij} \in \mathbb{R}^s$ such that such that $H_{ij}^{+} = \{\mu \in \Sigma : \hat{u}(\mu, a_i) \geq \hat{u}(\mu, a_j)\}$ and $H_{ij}^{-} = \{\mu \in \Sigma : \hat{u}(\mu, a_i) \leq \hat{u}(\mu, a_j)\}$.

Given Assumption H, we can represent the preferences by the set of normals $\{h_{ij}\}_{i,j}$.

Assumption H strengthens convexity of $\mathcal{A}_j^{\Delta}$ in the following sense. Convexity of $\mathcal{A}_j^{\Delta}$ imposes conditions only on the most preferred alternative in each state. Assumption H requires that for each pair of alternatives a_i and a_j, the simplex be separated into two convex sets - one where a_i is preferred and the other where a_j is preferred. For any σ, the set of states for which any two alternatives receive equal vote shares is characterized by a hyperplane, irrespective of the vote shares received by the other alternatives for these states. Assumption H imposes a similar structure on the preferences, requiring that the states where the voter is indifferent between any two alternatives lie on a hyperplane, irrespective of whether these alternatives are top-ranked or not in these indifferent states.

The next Lemma shows that assumptions E and H impose a particular linear dependence on the set of hyperplanes $\{h_{ij}\}$ through transitivity.

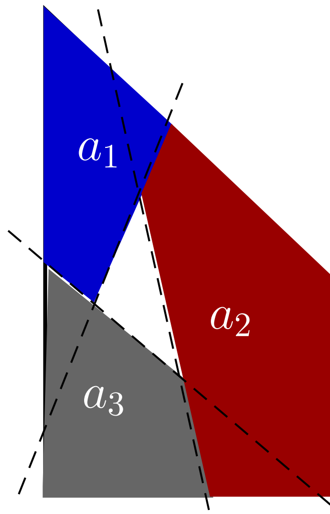

FIGURE 4. FIE failure with loops.

Lemma 3. *Suppose assumptions **M, E**, and **H** hold. Then, for any three alternatives $a_i, a_j, a_l \in A$, there exist positive constants α_{ij}, α_{jl}, and α_{il} such that*

$$\alpha_{ij} h_{ij} + \alpha_{jl} h_{jl} = \alpha_{il} h_{il}. \tag{3}$$

In words, for every three alternatives, the three hyperplanes that describe indifference between each pair are either parallel to each other or have a common intersection. In absence of assumption E, the hyperplanes could intersect pairwise outside the simplex. Thus, assumption E simply imposes a restriction on the preferences over distributions within the simplex, i.e., on preferences described by $v(\cdot)$.

Figure 4 makes the argument graphically by contradiction. Suppose there are three alternatives $\{a_1, a_2, a_3\}$ and suppose the dashed lines are the hyperplanes representing each pairwise separation (h_{12}, h_{13}, and h_{23}). and these do not intersect each other in the simplex. Note that the colored areas represent the region where an alternative is best for appropriately defined h_{12}, h_{13}, and h_{23}. It is easy to see that in the inner triangle generated by the pairwise lines we have a cycle over the preferences on the three alternatives.

Now, we are ready to state and prove the main theorem.

Theorem 3. *If Assumptions **M, E**, and **H** hold, there exists a strategy that achieves FIE.*

We have argued earlier that the challenge is to choose k strategy vectors to satisfy ${}_kC_2$ linear equations (" match" ${}_kC_2$ hyperplanes). The proof consists in showing that Lemma 3 imposes sufficient dependence among these ${}_kC_2$ equations so that we can guarantee a solution.

The main message of the section on rich state spaces is the following. For two or three alternatives, necessary and sufficient for information aggregation is convexity of the set of states for which a given alternative is top-ranked. For $k > 3$, the sufficient condition for FIE is that this convexity property must hold for every pairwise comparison of alternatives. This property has the flavor of Independence of Irrelevant Alternatives in Social Choice Theory. [[Is this too much? Yes, I do not think we should go that route]]We stress again that these conditions are satisfied only by special utility functions.

4.2.1. *Linear utility representation.* Before concluding this section, we present the following "representation theorem" for environments that allow FIE. This is a representation for utilities defined over probability distributions. More formally, denote the utility function over vectors on the simplex as $\tilde{u} : M \times A \to \mathbb{R}$ such that, $\tilde{u}(P(\cdot|\theta), a) := u(\theta, a)$, for all $a \in A$ and $\theta \in \Theta$. While this characterization is sharp when the state space is rich in the sense that $M = \Delta(X)$, the rich state assumption is not necessary for the characterization result to hold. We shall denote by $\mathcal{A}(\tilde{u})$ the partition of the state space induced by the utility function $\tilde{u}$. We will continue to assume that the signal space is $X = \{x_1, x_2, ..., x_s\}$.

Our result is the following. If the utility from each alternative is a linear function of the probability (or population proportion) of each signal, then the environment allows FIE. Conversely, for every environment that allows FIE, there exists a utility function linear in the probabilities which induces the same top-ranked alternative for each state.

Proposition 1. *Suppose there exists a utility $\tilde{u}(\mu, a) \equiv \tilde{u}_a \cdot \mu$, where $\tilde{u}_a \in \mathbb{R}^s$, $a \in A$, and all $\mu \in M \subseteq \Delta(X)$, then there exists a strategy that achieves FIE. Conversely, given some utility $v : M \times A \to \mathbb{R}$ suppose there exists a strategy that achieves FIE. Then, there exists $\tilde{u}'(\mu, a) = \tilde{u}_a \cdot \mu$, with $\tilde{u}'_a \in \mathbb{R}^s$, such that $\mathcal{A}(v) = \mathcal{A}(\tilde{u}')$ for all j.*

This characterization is also portrayed in example 3 and the discussion following it. An interpretation of the proposition (for a rich state space) is that the marginal change in utility from an alternative with respect to the proportion of any signal is independent of the proportion of the other signals. Alternatively, along the locus of indifference of any two alternatives, the rate at which the change in the proportion of one signal compensates for the change in proportion of another signal must be constant. In this sense, the tradeoff between any two signals should be unaffected by a third signal.

When the state space is rich, for every pair of alternatives a_i and a_j, the characterization allows us to develop a classification of signals in terms of which of the two alternatives is favored. Notice that this characterization depicts an environment that satisfies assumptions E and H. Also, the characterization is unique to the extent that, for all states μ lying on the hyperplane H_{ij} that describes indifference between a_i and a_j the vectors v_{a_i} and v_{a_j} satisfy

$$(v_{a_i} - v_{a_j}) \cdot \mu = 0.$$

Denote the r-th co-ordinate of the vector v_a by v_a^r. Now, define a partition $\{X_i, X_j, X_\phi\}$ of the signal space X in the following way: $X_i = \{x_r \in X \text{ if } v_{a_i}^r > v_{a_j}^r\}$, $X_j = \{x_r \in X \text{ if } v_{a_i}^r > v_{a_j}^r\}$ and $X_j = \{x_r \in X \text{ if } v_{a_i}^r = v_{a_j}^r\}$. Signals in X_i favor a_i and those in X_j favor a_j in the sense that higher proportion of any signals in X_i at the expense of any signal in X_j raises the utility difference between a_i and a_j. Moreover, since the strategy must have $\sigma_{ij}(x_r)$

in proportional to $v_{a_i} - v_{a_j}$, it must be the case that $\sigma_{ij}(x) > 0$ if $x \in X_i$ and $\sigma_{ij}(x) < 0$ if $x \in X_j$. In this sense, voting is sincere when FIE is achieved.

5. Equilibrium analysis

From the previous sections, it is clear that certain environments allow full information equivalence in the sense that there exist strategies that achieve FIE. However, it is not clear whether, even in such environments, voters have an incentive to use such strategies. In order to check whether voters find it in their interest to use such strategies, we consider voting as a game played in such environments. A game is defined as an environment $\{u, A, \Theta, X, P\}$ along with a number of players n. We fix an environment and consider a sequence of games by letting the number of voters grow. Following the logic in McLennan (1998), we show that under common preferences, any environment that allows FIE also has a sequence of Nash equilibrium profiles that achieves FIE.

Let us define the game G^n derived from the environment $\{u, A, \Theta, X, P\}$ along with a number of players n more formally. We will work directly with the mixed extension of the game. Each player's strategy set is $\Sigma = \{\sigma : \sigma = (\sigma_1, ..., \sigma_k), \sigma_j : X \to [0,1], \sum_j \sigma_j(x) = 1\}$, the set of all behavioral strategies. Endow Σ with the narrow topology so that it is a compact space. Let us abuse notation and use the letter a to denote a profile of choices: $a = (a^1, ..., a^n)$ where $a^i \in A = \{a_1, ..., a_k\}$ for each $i = 1, ..., n$. Let $u(\theta, a)$ be the utility at a pair (θ, a), that is, $u(\theta, a) = u(\theta, a_j)$, where a_j is the winner under the profile a.[9] Let $\sigma^{(n)} = (\sigma^1, ..., \sigma^n)$ denote a profile of behavioral strategies. At a state θ and profile $(x_1, ..., x_n)$ of signals, the (common) utility of a voter is $\sum_a \prod_{i=1}^n \sigma^i(a^i|x_i)u(\theta, a)$, where $\sigma^i(a^i|x_i) = \sigma^i_j(x_i)$ when $a^i = a_j$ (that is, when the choice of voter i at profile a is the alternative a_j.) Hence the common ex-ante utility at the profile $\sigma^{(n)}$ is

$$u(\sigma^{(n)}) = \int_\Theta \int_{X^{(n)}} \sum_a \prod_{i=1}^n \sigma^i(a^i|x_i)u(\theta, a) \otimes_{i=1}^n P(dx_i|\theta)P(d\theta)$$

where $X^{(n)}$ is the set of all profiles of signals $(x_1, ..., x_n)$. Observe that, for each θ, the term $\int_{X^{(n)}} \sum_a \prod_{i=1}^n \sigma^i(a^i|x_i)u(\theta, a) \otimes_{i=1}^n P(dx_i|\theta)$ is continuous in profiles of strategies $\sigma^{(n)}$ by the definition of the narrow topology on Σ and by virtue of the conditional independence of signals (that is, for each θ the "prior" $P(\cdot|\theta)$ is the product of marginals, so information is diffuse, and the expected utility is continuous in the product of the behavioral strategies – see Balder (1988)). Hence, by Lebesgue Dominated Convergence, $u(\sigma^{(n)})$ is continuous in $\sigma^{(n)}$. This ends the description of G^n.

Suppose that the profile $\sigma^{(n),*}$ is a maximizer of $u(\sigma^{(n)})$. The existence of such a maximizer follows from compactness of Σ and continuity of u in $\sigma^{(n)}$. Following McLennan (1998), $\sigma^{(n),*}$ is a Bayesian Nash equilibrium of the game G^n. It is straightforward to restrict to profiles of

[9] If there are ties, then view $u(\theta, a)$ as the expected utility of an unbiased tie-breaking rule. See the proof of Theorem 4 for a more explicit account of ties.

symmetric strategies and ensure existence of a symmetric BNE. The next theorem tells us that the sequence $\sigma^{(n),*}$ achieves FIE as long as the environment $\{u, A, \theta, X, P\}$ allows FIE.

Theorem 4. *If the environment (u, A, θ, X, P) allows FIE, there exists a sequence σ^n of Nash equilibria of the game G^n that achieves FIE., i.e., $W_n^{\sigma^n} \to 0$.*

6. Extensions and Discussion

6.1. **Genericity of Full Information Equivalence.** In this section, we formalize the idea that FIE is a generic property of finite state spaces as well as continuous and compact state spaces. In order to do so, we will rely on a recent result by Gizatulina and Hellwig (2015), which, translated to our environment immediately implies that the subset of $\mathcal{F}$ comprising of independent information structures is topologically generic. That is, such set is a *residual set*, which means that it contains the countable intersection of open and dense subsets of $\mathcal{F}$, under the appropriate topologies. Proposition 2 shows FIE is generic. By similar reasoning, FIE is generic for finite state spaces too, provided that the number of signals exceeds that of states. It must be mentioned here that all of the existing literature on information aggregation in large elections has concentrated on subcases of these two kinds of state spcaces. This includes Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1996), Myerson (1998), Wit (1998), Duggan and Martinelli (2005) and several others. While these have concentrated on binary state spaces, Feddersen and Pesendorfer (1997), McMurray (2014, 2017a) consider a unit interval state space, albeit with limited preference variation. Finally, McMurray (2017b) considers a state space that is a circle, but shows that it generically boils down to the unit interval. This result also tells us that the failure of FIE in example 2 was a non-generic case.

Let us define the appropriate topologies. Let $\Delta^\lambda(X)$ denote the subset of $\Delta(X)$ of probability measures that are absolutely continuous with respect to λ and have continuous densities. On $\Delta^\lambda(X)$, let us use the metric $d(f, \hat{f}) = \max_{x \in X} ||f(x) - \hat{f}(x)||$. And on the space of continuous mappings $C(\Theta, \Delta^\lambda(X))$, use the uniform topology, that is, the topology generated by the metric $\rho(P, \hat{P}) = \max_\theta d(f_{P(\cdot|\theta)}, f_{\hat{P}(\cdot|\theta)})$. Observe that the set of information structures is precisely such set of continuous functions, that is, $\mathcal{F} = C(\Theta, \Delta^\lambda(X))$. Let $\mathcal{F}^{ind} \subset \mathcal{F}$ denote the set of information structures satisfying independence. From Hellwig and Gizatulina (2015) we obtain:

Proposition 2. *$\mathcal{F}^{ind}$ is a residual subset of $\mathcal{F}$.*[10]

6.2. **Scoring rules.** While we have developed our conditions for FIE based on the plurality rule (and have not allowed abstention), our conditions hold for a much larger class of voting rules that are called scoring rules. These are rules where a voter can assign "scores" to each alternative, and the alternative with the highest score wins.

[10]The continuity assumption on the density f is used only for this result. In fact, the proof of Lemma 2 already uses only strong continuity of $\theta \mapsto P(\cdot|\theta)$, and all other results do not require any form of continuity of $\theta \mapsto P(\cdot|\theta)$.

We follow Myerson (2002) for defining a scoring rule. Let X be a finite set of signals and V be a positive integer. In an election with k alternatives, a *V-scoring rule* is a voting procedure where a ballot is a vector $v \in \mathcal{V} = \{0, ..., V\}^k$, and each element, v_j, is a number between 0 and V. We interpret v_j as the number of points the voter gives to alternative j. Ballots are aggregated by adding all points for every alternative, and the winner of the election is the alternative with most points. Under this framework, we can define several standard voting environments by imposing restrictions on the ballot. For example, in plurality voting, the voter is allowed to assign a single point to only one of the alternatives. In approval voting, the voter's ballot assigns one point to as many alternatives as she is willing to choose. We define two particular scoring rules below.

Definition 3 (Approval voting)**.** *An approval voting rule is equivalent to a* 1-scoring rule*.*

Definition 4 (Plurality voting)**.** $V = 1$ *and a ballot* $v \in \mathcal{V}$ *requires* $\sum_i v_i = 1$*.*

Note that we implicitly assume that plurality voting rules out abstention.

Let $\sigma_j^V : \mathcal{V} \times X \to [0,1]$, with $\sum_v \sigma_j^V(v,x) = 1$ for al $x \in X$ and all j, be a symmetric mixed strategy in a *V-scoring rule*, where $\sigma_j^V(v,x)$ is the probability that a player with signal x assigns $v \in \mathcal{V}$ points to alternative a_j. The next result says that FIE under plurality voting is necessary and sufficient for FIE under any scoring rule. The theorem also simultaneously establishes the equivalence of all V-scoring rules so far as the property of FIE is concerned.

Theorem 5. *Fix* $V \in \mathbb{N}_+$*. There exists a strategy that allows FIE in a* V-scoring rule *if and only if there is a strategy profile that achieves FIE in plurality voting.*

The above result says that a change in the voting rule from simple plurality to more general scoring rules (e.g. Borda count) fails to alter the set of environments that allow FIE: thus, if information is not aggregated under the plurality rule, it will not be aggregated under any scoring rule. Similarly, adding the option of abstention does not make a difference to the property of FIE.

The above result comes with two caveats. First, this equivalence holds only for large elections: for finite elections, there may well be a difference. In fact, Ahn and Oliveros (2016) shows that for any finite-sized election, the plurality rule performs the best among all scoring rules. Second, our result should not be taken to mean that scoring rules are irrelevant for large elections. The main import of Theorem 5 is that these rules matter only in a world where voters have non-common preferences.

Theorem 5 considers scoring rules that are symmetric across alternatives. This does not cover asymmetric rules like supermajority where one alternative must obtain a larger share of votes than the other alternative in order to be declared the winner. We define as a q-rule a voting rule where, among two alternatives a_1 and a_2, the former has to obtain at least $q \in (0,1)$ share of votes in order to win the election. The following proposition establishes that all q-rules are equivalent in terms of the set of environments that allow FIE.

Proposition 3. *Fix $q \in (0,1)$ and suppose $k = 2$. There exists a strategy that allows FIE in a* q-rule *if and only if there is a strategy profile that achieves FIE in plurality voting (i.e., $q = 0.5$).*

6.2.1. *Monotone Likelihood Ratio Property.* In our framework, we obtain conditions on FIE with general signal and state spaces. One way to see our result is to specialize our environment to ordered signal and state spaces. A standard informativeness assumption on signals in this setting is the Monotone Likelihood Ratio Property (MLRP), which ensures that a signal is a "sufficient statistic" of the state (Milgrom, 1981) in the sense that higher signals indicate higher states. Feddersen and Pesendorfer (1997) assume strict MLRP condition on signals and shows (albeit in a model of diverse preferences) that information is aggregated in equilibrium. We obtain a sufficient condition for an environment to allow FIE which entertains MLRP as a specific case. Let us restrict to the two alternative case, as the extension to multiple alternatives is immediate.

Definition 5 (Monotone Likelihood Ratio Property)**.** *Suppose $\Theta = [0,1]$ and $X = \{(x_1, ..., x_k) : (x_1, ..., x_k) \in [0,1]^k$, with $x_1 < x_2 < \cdots < x_k\}$. The signals are said to satisfy strict MLRP if, for any two signals $x < x'$, the likelihood ratio $\frac{P(x|\theta)}{P(x'|\theta)}$ is a decreasing function of θ.*

We obtain a sufficient condition for the existence of a strategy that achieves FIE in this environment that is weaker than MLRP. Assume that the prior P is non-atomic and has full support over $[0,1]$. Moreover, suppose that for some $\theta^* \in (0,1)$, a_1 is preferred for $\theta > \theta^*$ and a_2 is preferred for $\theta < \theta^*$. In other words, $\mathcal{A}_1 = (\theta^*, 1]$ and $\mathcal{A}_2 = [0, \theta^*)$.

Let $F(x|\theta) = \sum_{x_j \leq x} P(x_j|\theta)$ denote the cumulative distribution function of $P(\cdot|\theta)$. Strict MLRP implies that for every x, the cumulative distribution $F(x|\cdot)$ is a decreasing function. Now consider the following property: For each $\theta' \in \mathcal{A}_1$ and each $\theta'' \in \mathcal{A}_2$, we have for all $x \in X$

$$F(x|\theta') < F(x|\theta^*) < F(x|\theta'') \tag{4}$$

As long as the property (4) is satisfied, there exists a strategy that achieves FIE. To see that, let x^* be the smallest $x \in X$ such that $F(x|\theta^*) \leq \frac{1}{2}$. Now, set $\sigma(x) = 0$ for $x \leq x^*$ and $\sigma(x) = 1$ for $x > x^*$. It is easy to verify that the strategy profile σ achieves FIE.[11]

Note that the property (4) is weaker than strict MLRP. While strict MLRP implies that $F(x|\cdot)$ is decreasing over the entire interval $[0,1]$, property (4) does not require $F(x|\cdot)$ to be decreasing within $\mathcal{A}_1$ or within $\mathcal{A}_2$.

6.3. **Diverse Preferences.** So far we have assumed that all voters have the same preferences. In this section we extend our basic insight to a case where the voters in the electorate may have different preferences. We maintain the assumption that all voters are ex ante identical, and draw their information and preferences from some distribution conditional on

[11] When the state space is not $[0,1]$, a sufficient condition for obtaining an FIE strategy with signals $x_1 < ... < x_k$ is that (1) each $\theta^{piv} \in M^{piv}$ leads to the same cumulative distribution $F(x|\theta^{piv})$ for all $x \in X$, and (2) for any $\theta' \in \mathcal{A}_1$ and $\theta'' \in \mathcal{A}_2$, we obtain $F(x|\theta') < F(x|\theta^{piv}) < F(x|\theta'')$ for all $x \in X$.

the state. To do so, we retain the elements of the set-up in the main section and assume in addition that the private signal x is also payoff relevant. Thus, the private draw of an individual serves two functions: it is a view about the outcomes and it provides information about how others view the outcomes. We may think of $x_i = (s_i, t_i)$, where s_i is the common value component and t_i is the private value component of the preference. Notice that this is a general setting that can encompass many different environments. In particular, it admits the environments studied in Feddersen and Pesendorfer (1997) with continuous state space and Bhattacharya (2013) with just two states.

Consider, therefore, that voters preferences are given by $u : \Theta \times X \times A \to \mathbb{R}$. In this extension, we shall only consider two alternatives a_1 and a_2. We now normalize $u(\theta, x) = 1$ if $u(\theta, x, A) > u(\theta, x, B)$, $u(\theta, x) = 0$ if $u(\theta, x, A) < u(\theta, x, B)$, $u(\theta, x) = \frac{1}{2}$ if $u(\theta, x, A) = u(\theta, x, B)$. This normalization is innocuous for the feasibility result.

Since there is no commonly preferred candidate under diverse preferences, we have to be careful while defining the standard for information aggregation. We say that a strategy profile aggregates information if, for a large electorate, the outcome is the same as the outcome when the state is commonly known.

By the SLLN, asymptotically the proportion of voters that prefer a_1 to a_2, given a state θ, is $u(\theta) = \sum_{x \in X} u(\theta, x) P(x|\theta)$. In a large electorate, a_1 would get a vote share very close to $u(\theta)$ if the state were known to be θ. Therefore, under full information a_1 wins depending on whether $u(\theta)$ is greater or less than $\frac{1}{2}$. $\mathcal{A}_1 = \{\theta \in \Theta : u(\theta) > \frac{1}{2}\}$, $\mathcal{A}_2 = \{\theta \in \Theta : u(\theta) < \frac{1}{2}\}$, and $\mathcal{A}_{12} = \{\theta \in \Theta : u(\theta) = \frac{1}{2}.\}$. As before, we assume that the prior measure of the set $\mathcal{A}_{12}$ is zero. In the same vein as before, given a symmetric strategy profile σ, define $\mathcal{A}_i^\sigma$ as the set of states where alternative a_i wins almost surely in a large election.

We then say that an environment (u, Θ, X, P) allows FIE if there exists a strategy σ such that, for $i = 1, 2$ $P(\mathcal{A}_i \backslash \mathcal{A}_i^\sigma) = 0$. It is simple to verify now that the argument in the proof of Theorem 1 follows through line-by-line, so FIE is again characterized by the hyperplane condition. More formally, we have the following result,

Proposition 4. *Suppose $A = \{a_1, a_2\}$. An environment (u, Θ, X, P) and diverse preference allows FIE if and only if there exists a hyperplane H in $\Delta(X)$ such that $P(\cdot|\theta) \in H^+$ for $\theta \in \mathcal{A}_1$, $P(\cdot|\theta) \in H^-$ for $\theta \in \mathcal{A}_2$, and, if $M^{piv} \neq \emptyset$, $P(\cdot|\theta) \in H$ for $\theta \in M^{piv}$.*

Notice that since Feddersen and Pesendorfer (1997) result already tells us that information is aggregated in equilibrium, the existence of FIE strategies is trivial in their setting. More interestingly, while Bhattacharya (2013) concentrates on showing that, for any consequential rule, there exists an equilibrium that fails to aggregate information, it can be checked that in Bhattacharya's two-state setting, there always exists some feasible strategy that achieves FIE. It would therefore be very interesting to examine the conditions under which, in a general setting with diverse preferences, there exists some equilibrium sequence that aggregates information.

The proof of Theorem 4 explicitly utilizes the common value setting, and therefore does not automatically generalize to an environment with diverse preferences. In particular, we do not know the conditions under which the existence of a feasible strategy profile guaranteeing FIE also implies that FIE is achieved in equilibrium when there is preference diversity in the electorate. We believe that this is an important open question.

7. Conclusion

The existing literature on information aggregation in large elections has largely focussed on specific preference and information environments. In this paper, we consider general environments in order to analyze conditions under which information is aggregated. Preferences depend on the state of the world, and each state of the world is synonymous with a probability distribution over private signals. Therefore, preferences are simply mappings from allowable probability distributions over private information to rankings over the alternatives. In a large electorate, the frequency distribution over signals is approximately the same as the probability distribution. Thus, our question is whether the election achieves the outcome that would obtain if the entire profile of private signals were publicly known. If an environment permits a feasible strategy profile that can induce the full information outcome with a high probability in almost all states, we say that the environment allows Full Information Equivalence (FIE). Moreover, we are interested in whether such a strategy profile is incentive compatible, i.e., it constitutes a Nash equilibrium in the underlying game.

For most part of the paper, we consider the case where all voters would have agreed on their rankings if they had known the profile of signals. We provide a complete characterization of the environments that allow FIE when there are two alternatives. We provide separate necessary and sufficient conditions for environments to allow FIE for the general case of any number alternatives. We first consider finite state and signal spaces and show that as long as the number of possible signals is larger than the number of possible states, FIE holds generically. As a special case, whenever there are two states, FIE is always achieved in equilibrium. Then we turn to the case where the state space is rich in the sense that preferences are defined over all distributions on the simplex. In this case, FIE holds only for rather special environments. In other words, the property of FIE depends on relative richness of state and signal spaces: if the signal space is richer than the state space, then FIE is a generic property, but if the state space is richer than the signal space, then FIE is unlikely to obtain. Alternatively, if the information environment is sufficiently complex, there is no strategy profile that aggregates information. We like to stress here that the failure of FIE has nothing to do with equilibrium assessments over the states based on the criterion of one's vote being pivotal in deciding the election.

Under common preferences, as long as an environment allows FIE, there is a sequence of equilibria that achieves FIE. There may be other equilibrium sequences that do not aggregate information—but ours is only a possibility result. We also show that in the common preference environment, the voting rule does not matter for information aggregation: as long as FIE is achieved by the majority rule, FIE is achieved under a much larger class of voting rules.

In a later section, we consider diverse preferences, i.e., where the private information may signify both information about own preference and also about the distribution of preferences. In this case, we show that our two-alternative characterization goes through. However, feasibility of FIE does not automatically imply FIE in equilibrium when voting has the burden of aggregating information and preferences simultaneously. The negative results, however, go through as before: we identify the set of environments where FIE cannot be achieved. It would be interesting to find additional conditions under which feasibility of FIE indeed implies the existence of an equilibrium sequence that also achieves FIE in the case with diverse preferences.

Finally, one should note that we have not allowed communication between voters in our model. If communication were to be allowed in the case of common preferences, then everyone has an incentive to share their private information. Therefore, information would trivially be aggregated. In this context, our positive results are significant: we show that if the number of signals is larger than the number of states, then information aggregation does not require communication. On the other hand, in the case of diverse preferences, it is unclear whether truthful sharing of information is incentive compatible. It would be interesting to study the role of pre-voting deliberation in aggregating information when voters do not have common preferences.

8. References

(1) Acharya, Avidit and Adam Meirowitz. "Sincere Voting in Large Elections" (2017), Games and Economic Behavior, forthcoming

(2) Ahn, David and Santiago Oliveros. "Approval Voting and Scoring Rules with Common Values" Journal of Economic Theory (2016): 294-301

(3) Aliprantis, C. D., and K. C. Border. "Infinite Dimensional Analysis", Springer-Verlag. (2006).

(4) Austen-Smith, David, and Jeffrey S. Banks. "Information aggregation, rationality, and the Condorcet jury theorem." American Political Science Review (1996): 34-45.

(5) Balder, Erik, "Generalized Equilibrium Results for Games with Incomplete Information." Mathematics of Operations Research 13, (2013): 265-276.

(6) Berg, Sven. "Condorcet's jury theorem, dependency among jurors." Social Choice and Welfare 10, no. 1 (1993): 87-95.

(7) Bhattacharya, Sourav. "Preference Monotonicity and Information Aggregation in Elections." Econometrica 81, no. 3 (2013): 1229-1247

(8) Duggan, John, and César Martinelli. "A Bayesian model of voting in juries." Games and Economic Behavior 37, no. 2 (2001): 259-294.

(9) Feddersen, Timothy, and Wolfgang Pesendorfer. "Voting behavior and information aggregation in elections with private information." Econometrica: Journal of the Econometric Society (1997): 1029-1058.

(10) Gizatulina, Alia, and Martin Hellwig. "The Genericity of the McAfee-Reny Condition for Full Surplus Extraction in Models with a Continuum of Types."mimeo, Max Planck Institute for Research on Collective Goods.

(11) Ladha, Krishna K. "The Condorcet jury theorem, free speech, and correlated votes." American Journal of Political Science (1992): 617-634.

(12) Mandler, Michael. "The fragility of information aggregation in large elections." Games and Economic Behavior 74, no. 1 (2012): 257-268.

(13) McAfee, Preston, and Phillip Reny. "Correlated Information and Mechanism Design."Econometrica 60, no. 2 (1992): 395-421.

(14) McLennan, Andrew. "Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents." American Political Science Review (1998): 413-418.

(15) McMurray, Joseph. "Why the Political World is Flat: An Endogenous Left-Right Spectrum in Multidimensional Political Conflict", (2014) mimeo, Brigham Young University

(16) McMurray, Joseph. "Ideolgy as Opinion: A Spatial Model of Common Value Elections."American Economic Journal: Microeconomics, forthcoming

(17) McMurray, Joseph. "Voting as Communicating: Mandates, Multiple Candidates, and the Signaling Voter's Curse", (2017), Games and Economic Behavior: 199-223

(18) Milgrom, Paul R. "Good news and bad news: Representation theorems and applications." The Bell Journal of Economics (1981): 380-391.

(19) Myerson, Roger B. "Extended Poisson games and the Condorcet jury theorem." Games and Economic Behavior 25, no. 1 (1998): 111-131.

(20) Myerson, Roger B. "Comparison of Scoring Rules in Poisson Voting Games", Journal of Economic Theory (2002): 219-251

(21) Pesendorfer, Wolfgang, and Jeroen Swinkels "The Loser's Curse and Information Aggregation in Auctions,", Econometrica (1997): 1247:1281

(22) Siga, Lucas. "Complex Signals and Information Aggregation in Large Markets" mimeo, New York University (2016)

(23) Wit, Jörgen. "Rational choice and the Condorcet jury theorem." Games and Economic Behavior 22, no. 2 (1998): 364-376.

9. Appendix

9.1. **Proof of Lemma 1.**

Let $\{\hat{\sigma}_k\}_{k=1}^K$ be such that, for every x, we have

$$\begin{aligned}\hat{\sigma}_1(x) - \hat{\sigma}_2(x) &= h(x) - c_1^{-1}\\ \hat{\sigma}_2(x) - \hat{\sigma}_3(x) &= h(x) - c_2^{-1}\\ \ldots &= \ldots\\ \hat{\sigma}_{k-1}(x) - \hat{\sigma}_k(x) &= h(x) - c_{k-1}^{-1}\end{aligned}$$

Clearly, the $\hat{\sigma}$'s are bounded, so we can choose δ sufficiently large so that $\hat{\sigma}_j(x)+\delta \geq 0$ for all j and all x. By the same reason, we can choose $\varepsilon > 0$ sufficiently small such that $\sum_j(\hat{\sigma}_j(x)+\delta)\varepsilon \leq 1$ for every x. Let $R(x) = 1 - \sum_j(\hat{\sigma}_j(x)+\delta)\varepsilon$, and set $\sigma_j(x) = (\hat{\sigma}_j(x)+\delta)\varepsilon + R(x)/k$, so that $\sum_j \sigma_j(x) = 1$ for every x.

Then, $\sigma = (\sigma_1, ..., \sigma_k)$ is a well defined strategy. To show that σ achieves FIE, consider $\theta \in \Theta$ such that alternative j is best, that is, $\theta \in \mathcal{A}_j$. The mass of votes for alternative l is $\int \sigma_l(x)P(dx|\theta)$. We want to show that $\int(\sigma_j(x) - \sigma_l(x))P(dx|\theta) > 0$ for all $l \neq j$. First note that

$$\int(\hat{\sigma}_j(x) - \hat{\sigma}_l(x))P(dx|\theta) > 0 \Rightarrow \int(\hat{\sigma}_j(x) + \delta - \hat{\sigma}_l(x) - \delta)\varepsilon P(dx|\theta) > 0$$

$$\Rightarrow \int[(\hat{\sigma}_j(x)+\delta)\varepsilon + R(x) - [(\hat{\sigma}_l(x)+\delta)+R(x)]]P(dx|\theta) > 0 \Rightarrow \int(\sigma_j(x) - \sigma_l(x))P(dx|\theta) > 0.$$

Analogously

$$\int(\hat{\sigma}_j(x) - \hat{\sigma}_l(x))P(dx|\theta) < 0 \Rightarrow \int(\sigma_j(x) - \sigma_l(x))P(dx|\theta) < 0.$$

Consider $l > j$. By Property PS, $\int h(x)P(dx|\theta) > c_j^{-1}$, so $\int(h(x) - c_j^{-1})P(dx|\theta) > 0$. As $c_l > c_j$ for all $l > j$, we have $\int(h(x) - c_l^{-1})P(dx|\theta) > 0$ for all $l > j$. Then

$$\int(\hat{\sigma}_j(x) - \hat{\sigma}_l(x))P(dx|\theta) = \sum_{i=j}^{l}\int(h(x) - c_i^{-1})P(dx|\theta) > 0$$

and hence

$$\int(\sigma_j(x) - \sigma_l(x))P(dx|\theta) > 0.$$

Consider $l < j$. By Property PS, $\int h(x)P(dx|\theta) < c_{j-1}^{-1}$. So $\int(h(x) - c_{j-1}^{-1})P(dx|\theta) < 0$. As $c_l < c_j$ for all $l < j$, we have $\int(h(x) - c_{l-1}^{-1})P(dx|\theta) < 0$ for all $l < j$. Again, this means that

$$\int(\hat{\sigma}_l(x) - \hat{\sigma}_j(x))P(dx|\theta) = \sum_{i=l}^{j-1}\int(h(x) - c_i^{-1})P(dx|\theta) < 0$$

and hence

$$\int(\sigma_l(x) - \sigma_j(x))P(dx|\theta) < 0.$$

9.2. **Proof of Lemma 2.**

Consider a sequence of finite subsets Θ^m of Θ such that (i) $\Theta^m \subset \Theta^{m+1}$, (ii) $\Theta^m \to \Theta$ in Hausdorff sense, and (iii) the densities $f(\cdot|\theta)$ are independent, for all $\theta \in \Theta^m$. We can do this because P satisfies independence. In fact, if for any finite set $\{\theta_1, ..., \theta_L\}$ the densities $f(\cdot|\theta_\ell)$, $\ell = 1, ..., L$ were not independent, we would have $\sum_\ell f(x|\theta_\ell)\alpha_\ell = 0$ for every x with some of the weights α_ℓ being non-zero. Without loss, let $\alpha_1 \neq 0$. Then $\sum_{\ell=1}^{L} f(x|\theta_\ell)\hat{\alpha}_\ell = f(x|\theta_1)$, with $\hat{\alpha}_1 = \alpha_1 + 1$ and $\hat{\alpha}_\ell = \alpha_\ell$, for $\ell = 2, ..., L$. But then, setting ν to be $\sum_\ell \delta_\ell \hat{\alpha}_\ell$, where δ_ℓ is the point mass at θ_ℓ, we would have $\int f(x|\theta_\ell)\nu(d\theta) = f(x|\theta_1)$ for every x, which implies that $\int P(\cdot|\theta)\nu(d\theta) = P(\cdot|\theta_1)$, which in turn implies that $\hat{\alpha}_1 = 1$, or $\alpha_1 = 0$, and $\alpha_\ell = 0$ for $\ell = 2, ..., L$ by independence of P.

Fix a list $0 < c_0 < c_1 < c_2 < \cdots < c_{k-1} < c_k$. We want to show that property **PS** holds. Let $\hat{c}_i = c_i^{-1} + \varepsilon$, for $\varepsilon > 0$ smaller than the difference between any two c_i and c_j. By affine independence, for each m there is $h^m \in L^\infty(X)$ (in fact, we can choose h^m to have range in $[-1, 1]$) such that

$$\int h^m(x) f(x|\theta)dx = \hat{c}_i, \text{ for all } \theta \in \mathcal{A}_i^m,$$

where $\mathcal{A}_i^m = \Theta^m \cap \mathcal{A}_i$.

By Alaoglu's theorem (Aliprantis and Border (2006), Theorem 6.21), the so constructed sequence h^m has a weak*-convergent subsequence, so let h be its limit. By construction, for each $\theta \in \mathcal{A}_i^m$

$$\int h(x) f(x|\theta)dx = \hat{c}_i.$$

As $\mathcal{A}_i^m$ converges to $\mathcal{A}_i$, for each such θ we must also have $\int h(x)f(x|\theta)dx = \hat{c}_i$. Indeed, there must exist a sequence θ^m with $\theta^m \in \mathcal{A}_i^m$ such that $\theta^m \to \theta$; by strong continuity of $\theta \mapsto P(\cdot|\theta)$, $f(\cdot|\theta^m)$ norm-converges to $f(\cdot|\theta)$, so $\int h(x)f(x|\theta^m)dx \to \int h(x)f(x|\theta)dx$. Property **PS** is therefore verified.

9.3. **Proof of Lemma 3.**

Let $I = H_{ij} \cap H_{jm}$. Suppose first that I is non-empty. If $\mu \in I$ then $\hat{u}(\mu, a_i) = \hat{u}(\mu, a_j) = \hat{u}(\mu, a_m)$, hence $\mu \in H_{im}$. By assumption **H**, I is a vector space of dimension $k - 2$. It's orthogonal complement therefore has dimension 2. In this orthogonal complement, h_{ij} and h_{jm} lie on a two dimensional linear space, they are independent (because their corresponding hyperplanes intersect), and therefore, they span the space. Since h_{im} also lies in this space it must be true that there exists constants α_{ij}, α_{jm} such that $h_{im} = \alpha_{ij}h_{ij} + \alpha_{jm}h_{jm}$. Suppose instead I is empty. This is true whenever H_{ij} and H_{jm} are parallel. It is clear that there is always constants satisfying equation (3). Therefore, we establish that whether I is empty or not, there will always exist constants satisfying equation (3).

Next we show by contradiction that the constants need to be positive.

Case 1. Suppose $\alpha_{ij} < 0$, and $\alpha_{jm} < 0$. There are two possible scenarios: either there exists μ such that $\mu \cdot h_{ij} > 0$ and $\mu \cdot h_{jm} > 0$, or there exists μ such that $\mu \cdot h_{ij} < 0$ and $\mu \cdot h_{jm} < 0$. Otherwise, one of the alternatives is never the best choice. We will focus on the former and note that the analogous argument holds inverting the inequalities. Thus, $\hat{u}(\mu, a_i) > \hat{u}(\mu, a_j)$ and $\hat{u}(\mu, a_j) > \hat{u}(\mu, a_m)$. By transitivity, $\hat{u}(\mu, a_i) > \hat{u}(\mu, a_m)$. On the other hand, $\mu \cdot (\alpha_{ij} h_{ij} + \alpha_{jm} h_{jm}) < 0$. Thus $\mu \cdot h_{im} < 0$, and then, $\hat{u}(\mu, a_i) < \hat{u}(\mu, a_m)$, a contradiction.

Case 2. Suppose $\alpha_{ij} > 0$, and $\alpha_{jm} < 0$. Consider μ such that $\mu \cdot h_{im} = 0$ and $\mu \cdot h_{ij} \neq 0$. $\mu \cdot h_{im} = 0$ implies $\mu \cdot (\alpha_{ij} h_{ij} + \alpha_{jm} h_{jm}) = 0$. Then $\alpha_{ij} \mu \cdot h_{ij} = -\alpha_{jm} \mu \cdot h_{jm}$. There are two possibilities: (i) $\mu \cdot h_{ij} > 0$. This implies $\mu \cdot h_{jm} > 0$. Then $\hat{u}(\mu, a_i) > \hat{u}(\mu, a_j) > \hat{u}(\mu, a_m)$. (ii) $\mu \cdot h_{ij} < 0$. This implies $\mu \cdot h_{jm} < 0$. Then $\hat{u}(\mu, a_i) < \hat{u}(\mu, a_j) < \hat{u}(\mu, a_m)$. In either case this is a contradiction since $\mu \cdot h_{im} = 0$ implies that $\hat{u}(\mu, a_i) = \hat{u}(\mu, a_m)$.

Case 3. Suppose $\alpha_{ij} < 0$, and $\alpha_{jm} > 0$. This is symmetric to the case 2.

Finally, let $\alpha_{im} = 1$ and the result follows.

9.4. **Proof of Theorem 1.**

"If" direction. Let $h(x)$ be the norm of a hyperplane with the separating properties prescribed by the theorem. Let $\underline{c} \equiv \inf_{\theta \in \Theta} \int -h(x) P(dx|\theta)$, and $\bar{c} \equiv \sup_{\theta \in \Theta} \int -h(x) P(dx|\theta)$ and define $c_2^{-1} = \varepsilon$, $c_1^{-1} = |\underline{c}| + \varepsilon$, and $c_0^{-1} = \bar{c} + |\underline{c}| + \varepsilon$ for a small $\varepsilon > 0$. Then, property **PS** is satisfied for $0 < c_0 < c_1 < c_2$ and by Lemma 1, there exists a strategy that achieves FIE.

"Only if" direction. Consider a strategy σ that achieves FIE. Define $h(x) = \sigma_1(x) - \sigma_2(x)$ and let $H = \{\mu \in \Delta(X) : \int_{x \in X} h(x) \mu(dx) = 0\}$, then it is immediate to note that $P(\cdot|\theta) \in H^+$ for $\theta \in \mathcal{A}_1$, $P(\cdot|\theta) \in H^-$ for $\theta \in \mathcal{A}_2$, and, if $M^{piv} \neq \emptyset$, $P(\cdot|\theta) \in H$ for $\theta \in M^{piv}$.

9.5. **Proof for Example 1.**

In order for a_1 to win almost surely in both L and R, we require the strategy $\sigma(\cdot)$ to satisfy the following two conditions respectively

$$\begin{aligned} z_{12}^{\sigma}(L) &= p\sigma_{12}(x) + (1-p)\sigma_{12}(y) > 0 \\ z_{12}^{\sigma}(R) &= (1-p)\sigma_{12}(x) + p\sigma_{12}(y) > 0 \end{aligned}$$

Taken together, we must have $\sigma_{12}(x) + \sigma_{12}(y) > 0$, violating the condition for a_2 winning in state M for large n, given by

$$z_{12}^{\sigma}(M) = \frac{1}{2}\sigma_{12}(x) + \frac{1}{2}\sigma_{12}(y) < 0$$

and the result follows.

9.6. **Proof for Example 2.**

For any strategy $\sigma = (\sigma_1(\cdot), \sigma_2(\cdot))$, the vote share difference function is given by

$$z_{12}^{\sigma}(\theta) = \theta\sigma_{12}(x) + (1-\theta)\sigma_{12}(y)$$

Notice that this function is continuous and linear in θ.

For σ to satisfy FIE in the first environment, we must have $z_{12}^{\sigma}(\theta) > 0$ for $\theta > t$ and $z_1^{\sigma}(\theta) < 0$ for $\theta < t$. By continuity, we must have $z_{12}^{\sigma}(t) = 0$. Any σ that satisfies (i) $z_{12}^{\sigma}(t) = 0$ and (ii) $z_{12}^{\sigma}(\cdot)$ is an increasing function leads to FIE. It is easy to check that we can always find some σ with these properties.

For FIE in the second environment, we must have $z^{\sigma}_{12}(\theta) > 0$ for $\theta \in (t_1, t_2)$ and $z^{\sigma}_{12}(\theta) < 0$ for $\theta \in [0, t_1) \cup (t_2, 1]$. However, since $z^{\sigma}_{12}(\cdot)$ is linear in θ for every strategy σ, there is no symmetric strategy profile that achieves FIE in this case.

9.7. **Proof for Example 4.**

The result follows by noting that there exists no hyperplane that can separate $\mathcal{A}_1$ from $\mathcal{A}_2$ (i.e., $\mathcal{A}_1$ is a circle enclosed by $\mathcal{A}_2$) and then applying Theorem 1.

9.8. **Proof of Theorem 3.**

Consider the system,

$$\alpha_{ij} h_{ij} + \alpha_{jl} h_{jl} = \alpha_{il} h_{il} \text{ for all } i, j, l \tag{5}$$

Lemma [[3]] guarantees that the equations is well defined. The number of equations in the system (5) is given by ${}_kC_3$. Notice that h_{ij} are parameters of the equation given by the preferences. We will show that the system in (5) has a non trivial solution for the variables α's. Furthermore, if the solution is non trivial, all α's are strictly positive. To show this last assertion, suppose instead that there exists some $\alpha_{ij} = 0$, then $h_{jl} = c.h_{il}$, for some constant c. [[However, $H_{jl} = H_{il} \neq H_{ij}$.]] This is not possible because the first equality implies that there exist some μ such that $\hat{u}(\mu, a_i) = \hat{u}(\mu, a_j) = \hat{u}(\mu, a_l)$ but the second inequality implies that $\hat{u}(\mu, a_i) \neq \hat{u}(\mu, a_j)$ for all μ.

The number of variables α's is ${}_kC_2$. For $k < 6$, ${}_kC_2 > {}_kC_3$ and therefore the system has a non trivial solution. However, for $k \geq 6$, there are more equations than unknowns. We need to show that there are sufficiently many linearly dependent equations so that the system has a solution.

Consider the subsystem of equations in which we fix an alternative, that without loss of generality we will call alternative 1, and we combine with all the other possible combinations of the remaining two alternatives. This is the set of equations containing all equations in which alternative 1 is present. The number of equations in this subsystem is given by ${}_{(n-1)}C_2 < {}_nC_2$ and contains all α's, and therefore it has a non trivial solution. It only remains to show that any equation in the system given by (5) can be generated using this subsystem. For simplicity of exposition, and without loss of generality, consider an equation with alternatives $(2, 3, 4)$:

$$\alpha_{23} h_{23} + \alpha_{34} h_{34} = \alpha_{24} h_{24} \tag{6}$$

This equation will not be contained in our subsystem because alternative 1 is not present. Consider the following three equations from our subsystem:

$$\alpha_{12} h_{12} + \alpha_{23} h_{23} = \alpha_{13} h_{13} \tag{7}$$

$$\alpha_{12} h_{12} + \alpha_{24} h_{24} = \alpha_{14} h_{14} \tag{8}$$

$$\alpha_{13} h_{13} + \alpha_{34} h_{24} = \alpha_{14} h_{14} \tag{9}$$

We do the following operation: equations (7) minus equation (8) plus equation (9) and we note that this is equal to equation (6). Since the choice of alternatives is without loss of generality we convince ourselves that our subsystem generates the full system.

Let $\Upsilon = \{h_{1j}\}_{j>1}$. For all $h_{1j} \in \Upsilon$ let τ_1, and τ_j be such that $\tau_1 - \tau_j = \alpha_{1j} h_{1j}$. There are n variables τ's and $n-1$ equations so this system has a solution.

Finally, for all i, choose $\sigma_i = (\tau_i + \delta)\epsilon$, where δ is a sufficiently large constant such that $\tau_i + \delta \geq 0$, and $\epsilon = (\mathbf{e} \cdot \sum_i (\tau_i + \delta))^{-1}$. Then, $\sigma_i \geq 0$, and $\sum_i \sigma_i = \mathbf{e}$, and therefore constituting a well defined symmetric mixed strategy profile.

To show that this strategy aggregates information we need to show that if alternative a_i is the best, then the strategy selects alternative a_i over a_j, for any $a_j \in A$.

Consider the simplest case where alternative 1 is the best alternative. By construction, for all $j \neq 1$, $u(\mu, a_1) > u(\mu, a_j) \iff \mu \cdot h_{1j} > 0 \iff \mu \cdot (\tau_1 - \tau_j) > 0 \iff \mu \cdot (\sigma_1 - \sigma_j) > 0$. Then, alternative 1 obtains more votes than alternative j. The same relationship holds with weak inequality and equality.

Consider the case where alternative $a_i \neq 1$ is the best alternative. Then, $u(\mu, a_i) > u(\mu, a_j) \iff \mu \cdot h_{ij} > 0 \iff \mu \cdot (\alpha_{1j} h_{1j} - \alpha_{1i} h_{1i}) > 0 \iff \mu \cdot (\tau_1 - \tau_j - \tau_1 + \tau_i) = \mu \cdot (\tau_i - \tau_j) \iff \mu \cdot (\sigma_i - \sigma_j) > 0$.

Therefore, the strategy chooses the right candidate always in the limit and this concludes the proof.

9.9. **Proof of Proposition 1.**

We can assume that $\tilde{u}_a \in \Delta(X)$ without loss of generality because a positive affine transformation of v_a always lies in $\Delta(X)$. Choose the strategy $\sigma_a = \tilde{u}_a$ with the interpretation that $\sigma_a(x_\ell)$ is the ℓ-th co-ordinate of v_a. Then, we have $z_a^\sigma(\mu) = \tilde{u}_a \cdot \mu = \tilde{u}(\mu, a)$ for all $\mu \in \Delta(X)$, which guarantees FIE. For the converse, simply let $\tilde{u}'_a := \sigma_a$ for every alternative a.

9.10. **Proof of Theorem 4.**

Recall that, for a given symmetric profile of strategies σ, $p_n^\sigma(y|\theta)$ denotes the probability of a vector of proportions y, given θ. The definition readily extends to asymmetric profiles $\sigma^{(n)}$. The probability that an alternative a_j wins the election given $\sigma^{(n)}$ and θ, denoted by $q_n^{\sigma^{(n)}}(a_j|\theta)$, is then

$$q_n^{\sigma^{(n)}}(a_j|\theta) = \sum_{y \in E_n^0} p_n^{\sigma^{(n)}}(y|\theta) + \sum_{m=1}^{k-1} \sum_{y \in E_n^m} \frac{1}{m+1} p_n^{\sigma^{(n)}}(y|\theta)$$

where E_n^0 is the set of proportions y where $y_j > y_i$ for all $i \neq j$ and E_n^m is the set of proportions y where $y_j = y_i > y_\ell$ for all $\ell \neq i, j$ and for exactly m indices i. In words, E_n^0 is the set where a_j gets strictly more votes than all other alternatives and E_n^m is the set where a_j is tied at the top with exactly m other alternatives, in which case a_j wins with probability $\frac{1}{m+1}$. Observe that, with such definition in hands, we can write $u(\sigma^{(n)})$ as

$$u(\sigma^{(n)}) = \int_\Theta \sum_{j=1}^{k} u(\theta, a_j) q_n^{\sigma^{(n)}}(a_j|\theta) P(d\theta).$$

For each size n of electorate, consider a symmetric profile of strategies σ (recall our notation that σ without a superscript denotes both a single strategy and a profile where each voter uses the same strategy). For each θ, the proportion of votes for a_j converges to $z_j^\sigma(\theta)$ with $P(\cdot|\theta)$-probability one as $n \to \infty$. Hence, $q_n^\sigma(a_j|\theta)$ converges for each θ, so Lebesgue Dominated Convergence implies that $u(\sigma^\infty) = \lim_{n\to\infty} u(\sigma^n)$ is well defined.

Observe that if the symmetric profile $\hat{\sigma}^\infty$ achieves FIE, then $u(\hat{\sigma}^\infty)$ is the maximum attainable value: for P-almost every $\theta \in \mathcal{A}_j$, the alternative a_j wins. States in $\mathcal{A}_i \cap \mathcal{A}_j$ are irrelevant for the evaluation above because they are of P-measure zero. So, given that $u(\sigma^n)$ is linear in $u(\theta, a_j)$, the claim is verified.

For each finite electorate $\{1, ..., n\}$, choose $\sigma^{(n)}$ as a maximizer of $u(\sigma^{(n)})$. We know such profile is an equilibrium of the corresponding game G^n. We also know that $u(\hat{\sigma}^\infty)$ is the maximum feasible value of the ex-ante utility. Hence

$$u(\hat{\sigma}^\infty) \geq u(\sigma^\infty) = \lim_n u(\sigma^n) \geq \lim_n u(\hat{\sigma}^n) = u(\hat{\sigma}^\infty),$$

establishing the result. In fact, if $W_n^{\sigma^n}$ were not to converge to zero, then we would have to have, say, $m(\mathcal{A}_j \backslash \mathcal{A}_j^{\sigma^\infty}) > 0$ for some j. That is, a set of positive measure in $\mathcal{A}_j$ where an alternative $a_i \neq a_j$ wins under σ^∞, whereas we know that no such set exists for $\hat{\sigma}^\infty$. But then $u(\hat{\sigma}^\infty) > u(\sigma^\infty)$, contradicting what we just established.

9.11. **Proof of Proposition 3.**

Suppose, an environment allows FIE for some $q \in (0,1)$ and let $\sigma(\cdot)$ be the strategy that achieves FIE, with the interpretation that $\sigma(x)$ is the probability of voting for a_1 given signal $x \in X$. Now consider any other $q' \in (0,1)$. Replacing $\sigma(\cdot)$ by $\sigma'(\cdot) = q' + \epsilon(\sigma(\cdot) - q)$, we can ensure that the strategy σ' achieves FIE given voting rule q'. We make ϵ small enough to ensure $\sigma'(\cdot)$ is a valid strategy function.

9.12. **Proof of Theorem 5.**

Fix $V \in \mathbb{N}_+$. The *if* direction holds trivially: the strategy set under plurality voting is a subset of the strategy set under a *V-voting rule*. For the *only if* direction, suppose there exists a strategy profile σ that allows FIE. Let $\sigma_j^{\text{sum}}(x) = \sum_{v \in \mathcal{V}} v_j \sigma^V(v, x)$ be the expected number of points assigned to alternative j by a voter with signal x, and $\sigma^{\text{sum}}(x) = (\sigma_1^{\text{sum}}(x), ..., \sigma_k^{\text{sum}}(x))$.

Choose $\varepsilon > 0$ sufficiently small such that $\sum_{j=1}^k \varepsilon \sigma_j^{\text{sum}}(x) \leq 1$, for all $x \in X$. Define $R(x) = 1 - \sum_{j=1}^k \epsilon \sigma_j^{\text{sum}}(x)$, and let $\sigma_j^{\text{PV}}(x) = \epsilon \sigma_j^{\text{sum}}(x) + \frac{R(x)}{k}$. We want to show that $\sigma_j^{\text{PV}}(x)$ is a well defined plurality voting rule and that it chooses the same alternative as σ^V for all θ almost surely for n sufficiently large. By construction, $\sigma_j^{\text{PV}}(x) \in [0,1]$, and for all x,

$$\sum_j \sigma_j^{\text{PV}}(x) = \sum_{j=1}^k \left(\epsilon \sigma_j^{\text{sum}}(x) + \frac{1 - \sum_{l=1}^k \epsilon \sigma_l^{\text{sum}}(x)}{k} \right) = \sum_j \epsilon \sigma_j^{\text{sum}}(x) + 1 - 1 \sum_j \epsilon \sigma_j^{\text{sum}}(x) = 1.$$

Next we show that plurality voting chooses the same alternative as the *V-scoring rule*. First, note that in state θ the expected number of points received by alternative j is given by

$$\sum_{x\in X} \sigma_j^{\text{sum}}(x)P(x|\theta).$$

Then, as the population grows large, the difference in votes between alternative i and j, given θ under *V-scoring rule* is given by

$$\sum_{x\in X} \sigma_i^{\text{sum}}(x)P(x|\theta) - \sum_{x\in X} \sigma_j^{\text{sum}}(x)P(x|\theta)$$

Since $\sigma_j^{\text{PV}}(x)$ is an afine transformation of $\sigma_i^{\text{sum}}(x)$ the above difference is positive if and only if the following difference is positive:

$$\sum_{x\in X} \sigma_i^{\text{PV}}(x)P(x|\theta) - \sum_{x\in X} \sigma_j^{\text{PV}}(x)P(x|\theta).$$

This later expression is the expected difference in votes between alternative i and j. For n large, if i wins in a *V-scoring rule*, i wins in plurality voting and the result follows.

Department of Economics, University of Rochester, Email:paulo.barelli@rochester.edu.

Department of Economics, Royal Holloway University of London, Email:sourav.bhattacharya@rhul.ac.uk.

Department of Economics, New York University Abu Dhabi, Email: lucas.siga@nyu.edu.