Full Information Equivalence in Large Elections

Paulo Barelli* Sourav Bhattacharya† Lucas Siga‡

Preliminary - please do not circulate
December 21, 2017

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. We provide general conditions on the environment ensuring existence of a sequence of equilibria of the voting game that efficiently aggregates information as the population size grows to infinity. The conditions explore the geometry of partitions on the distributions over private signals induced by the common state-dependent utility of the voters. Such conditions are met generically when the signal space is rich enough relative to the state space, and fail robustly when the state space is rich relative to the signal space.

1 Introduction

In an election, voter preferences over candidates depend on a myriad of factors like their policy positions on various important issues, their past voting history, party affiliation, the state of the economy, geopolitical situation and so forth. Different voters are likely to hold different information about aspects of the underlying situation that matter for the voting decision. Moreover, such information is invariably noisy. We ask the classic question of information aggregation in elections: does the electoral outcome based on individual votes reflect all the information dispersed in the electorate? We analyze this question under the assumption that voters have the same underlying preference over the electoral alternatives, and any difference in induced preference is due to different information. Under this assumption, our question is equivalent to whether elections are guaranteed to choose the best candidate.

In contrast to the existing literature on the topic that analyzes this issue in very specific environments, we provide a novel way of analyzing the problem which enables us to say whether information can or cannot be aggregated in general preference and information environments.

*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.
‡Division of Social Sciences, New York University Abu Dhabi, Email: lucas.siga@nyu.edu.

The existing literature follows an insight by Condorcet (1786), known today as Condorcet Jury Theorem (CJT): if in a large two-candidate election, each voter votes for the correct candidate independently with probability $p > \frac{1}{2}$, then the majority is almost surely correct by the Law of Large numbers. Subsequent work (e.g., Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1997), Wit (1998), Myerson (1998), Duggan and Martinelli (2005)) generalizes this result in game-theoretic frameworks. These papers aim to demonstrate the following: if voters' preferences depend on an uncertain state variable and they receive noisy information about the realized state, then there exists a sequence of Nash equilibria with the property that the most preferred candidate wins almost surely in every state. Since the outcome is as if the state (ergo, the profile of private signals) were common knowledge, elections are then said to be *full information equivalent* (FIE). So far, the common practice in the literature is to assume that there are two states and two signals: each state is assumed to capture all situations under which a given candidate is optimal, and therefore a signal is a noisy opinion of which candidate is the better choice.[1]

Our work takes a different approach. We recognize that voter preferences depend on many different issues and similarly, different voters have private information about different aspects that matter for the decision. We therefore allow the state and signal space to be very general and look for conditions on the environment for which there exists a *feasible* strategy profile that achieves FIE in a given environment. We draw upon an insight from McLennan (1998) to argue that if there is a feasible strategy profile that achieves FIE, then there is also some equilibrium strategy profile that aggregates information. Hence the set of environments where FIE is achieved in (some) equilibrium is exactly same as the set of environments where FIE is achieved by some feasible strategy profile. An implication is that, in common-value environments, the property of information aggregation has nothing to do with equilibrium inferences conditioning on pivotality. For up to three alternatives, we obtain a complete characterization of environments where FIE is feasible. When there are more than three alternatives in contention, we provide general sufficient conditions.

Our conditions explore the geometry of partitions of the simplex of distributions over private signals induced by the state-dependent utility function and the information environment. In particular, FIE is related to whether or not such partitions consist of convex polytopes with facets defined by hyperplanes satisfying some properties. Convexity is the most important piece here: roughly speaking, it provides restrictions on the heterogeneity of distributions over private signals arising from states where one particular alternative is socially preferred. This can be viewed as a very general identifiability condition allowing one to infer information over states from observation of signals. In fact, one sufficient condition for FIE is exactly a rank condition on the family of conditional distributions over signals

[1] Feddersen and Pesendorfer (1997) and McMurray (2017) go beyond the binary formulation and assume an ordered state space which can be interpreted to be representing a single-issue election.

given states. An immediate corollary is that, in the case of finitely many states and signals, FIE is generic whenever we have more signals than states – in this case, a generic matrix of conditional distributions must have full rank. In the general case, however, one cannot generically guarantee linear independence of distributions, and neither can one generically guarantee convexity of the induced partition.

A special case of our positive result in discrete environments is that information is *always* aggregated whenever there are two alternatives, two states, and at least two signals. Hence, a take-away from our analysis is that it is misleading to lump into a single state all situations for which a given alternative is optimal, as has been done by much of the previous literature.

Before going to the formal conditions, we present a few simple examples to illustrate how information aggregation may fail. The first pair of examples is about single-issue politics and the second example is about multi-issue elections.

Consider a 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 frequently for more left-leaning positions and an x'-signal arises more frequently for more right-leaning positions. Since the private signal only tells 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.[2] Observe that the induced partition on distributions over signals x and x' is not convex: distributions "in the middle", with enough weight on both x and x' are associated with moderate policies and extreme distributions are associated with extreme policies, and this latter set is obviously not convex.

The next example illustrates the difficulty of aggregating information in multi-issue settings. Suppose there are three possible signal realizations: x, x' and x'', and assume that preferences are defined over the population proportion of each signal, so that we have infinitely many states: one state for each vector of proportions. In this case, the only utility functions that achieve FIE are those which are linear in the proportion of each signal. For the sake of concreteness, consider a country voting in a referendum on whether to stay or leave a politico-economic union (e.g., the "Brexit" vote in May 2016). A central tradeoff that drives voter preferences is that trade induces growth but leads to immigration as well, leading to

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

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. Now, assume that each voter receives a signal about exactly one of these three factors, and the frequency of a signal in the population depends on the strength of the factor. Voter preferences depend on the proportion of each type of signal in the population: if the proportion of signals about growth is high enough then voters prefer to stay, and if there is a very high proportion of signals about immigration is large then they prefer to exit. Our results say that information is aggregated if and only if the net utility from the exit option is linear in the proportion of each signal: in other words, if the tradeoff between immigration and growth is deemed to be independent of how much the immigrants contribute to the economy. 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, except for very special situations. In more specific terms, we see that the partition over distributions over signals induced by states is bound to not be made of convex sets except for very special cases.

While our set-up is very standard except for the generality, we take a distinct approach that allows for a convenient geometric representation. We focus on the implications of preferences over conditional distributions no signals arising in the states. Each state is thus mapped to a vector on the simplex (the space of all probability distributions) over the signals, and each such vector is associated with the corresponding ranking over alternatives. Notice that the expected vote share for any 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. As a result, the set of probability vectors for which an alternative obtains a fixed vote share is a hyperplane on the simplex over signals. This allows us to express all the conditions for FIE using separating hyperplanes on the simplex.

To describe our results, suppose that there are k alternatives $\{a_1, ..., a_k\}$, $k \geq 2$, and s signals $\{x_1, ..., x_s\}$, $s \geq 2$. Denote the set of conditional distributions over signals arising in states where a_i is the best choice by $\mathcal{A}_i^\Delta$. We first deal with the case of up to three alternatives (Section 3). Theorem 1 states that there exists a strategy that achieves FIE if and only if $\mathcal{A}_1^\Delta$, $\mathcal{A}_2^\Delta$ and $\mathcal{A}_3^\Delta$ can be separated by what we call a "restricted 3-partition" of the simplex: a partition defined by at most three distinct hyperplanes.

Theorem 1 demonstrates a fundamental difficulty in aggregating information when the state space is infinite. Consider the case of two alternatives. FIE requires that the two alternatives must obtain exactly equal vote shares for all pivotal states, that is, states around which the ranking flips. Since the vote share functions are linear in the vectors on the simplex, it must be the case that all pivotal states must lie on a hyperplane. In other words, a small perturbation in preferences around the pivotal states will lead to a violation of FIE. In the special case when preferences are defined over the entire simplex, FIE requires that $\mathcal{A}_1^\Delta$ and

$\mathcal{A}_2^\Delta$ form a convex partition of the simplex.

When the state space is discrete, there might be no pivotal states and consequently we have more freedom in choosing a strategy that induces the appropriate separation. For example, suppose there are only two states and two alternatives. Since $\mathcal{A}_1^\Delta$ and $\mathcal{A}_2^\Delta$ are two singletons, they can always be separated by a hyperplane. This tells us that the two-state formulation makes information aggregation trivial by imposing the restriction that there is exactly one distribution of signals for which a given alternative is optimal. Next, consider the case of three states and at least three signals. It is easy to check that we can separate any two vectors on the simplex from the third one by a hyperplane, as long as the three vectors are not collinear.

In section (Section 4), we consider the general case of $k > 3$ alternatives and show that the contrast between discrete and continuous environments in the $k \leq 3$ case is a sharp illustration of a more general phenomenon. We still have the same necessary condition for FIE: each pair $\mathcal{A}_i^\Delta$ and $\mathcal{A}_j^\Delta$ must be separated by a hyperplane. Notice that this condition implies that if there is a set of pivotal states in the neighborhood of which the top-ranked alternative changes from a_i to a_j, the respective conditionals must lie on a hyperplane. However, the necessary condition is not always sufficient, and we provide two sets of sufficient conditions for FIE.

Theorem 2 strengthens the necessary condition by imposing a restriction on the family of hyperplanes that separate the sets $\mathcal{A}_i^\Delta$. In a nutshell, when such hyperplanes are parallel to one another, FIE can be achieved. This result was originally developed in the context of information aggregation in auctions by Mihm and Siga (2017). While such parallel separation seems a very strong condition, it can be ensured whenever a rank condition is met. More precisely, Lemma 2 shows that we can obtain separation with parallel hyperplanes if the set of conditional probability vectors satisfies independence in addition to some regularity conditions. Therefore, a sufficient condition for FIE in general environments is linear independence (Corollary 2). For discrete environments, linear independence holds generically whenever there are at least as many signals as states.

When there are more states than signals, linear independence fails and so we cannot hope for separation with parallel hyperplanes. We are led to a more refined form of separation, which we refer to as "star-shaped separation: roughly, the hyperplanes are allowed to not only be not parallel, but also to intersect with one another in some specific ways. More precisely, Theorem 3 establishes that FIE can be achieved if for *every pair* of alternatives a_i and a_j, a hyperplane on the simplex separates the vectors for which a_i is preferred over a_j from those for which a_j is preferred over a_i.[3]

[3]Strictly speaking, Theorem 3 also requires a richness condition and an extension condition. Richness is imposed only for convenience of the analysis – the general idea carries through without richness. Extension allows us to focus on the relatively simple condition described above, without having to impose that if

When the state space is rich, the utility function can be thought to be defined directly on the probability distributions over signals. Proposition 1 provides a condition on this utility function which is equivalent to the pairwise separation condition in Theorem 3. The Proposition 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 conversely (ii) for any environment that allows FIE, there exists another environment with the same top-ranked alternative for each probability vector which admits a linear utility representation. While the linear utility representation is not a complete characterization of environments that allow FIE, it does indeed provide a tight characterization of how the simplex is partitioned into the sets $\mathcal{A}_j^{\Delta}$ in environments that allow FIE. The linearity result highlights the problem of aggregating information when preferences are defined simply over distributions of private signals. When there are more than two possible signals, FIE occurs only if the "marginal rate of substitution" between proportions of any two signals is independent of the proportion of every other signal.

Section 5 establishes that feasibility translates into FIE in equilibrium. Formally, whenever an environment allows a strategy that achieves FIE, there is a sequence of Nash equilibria in the same environment such that as the number of voters grows unboundedly, the ex-ante probability of the correct alternative being chosen converges to 1 (Theorem 4). This is basically the same result as in McLennan (1998), except that we can relax McLennan's requirement that feasibility be in symmetric strategies since we are only dealing with large electorates. Our contribution over McLennan's is the identification of environments where FIE can and cannot be achieved in the limit.

Finally, in Section 6, we provide two important extensions of our model to show that our results hold in a wide range of settings that have been studied in the literature. Up to that point, we obtain all our conditions using simple plurality rule (the alternative obtaining the highest number of votes is the winner). Theorem 5 says that a change in the voting rule to any other scoring rule (e.g. approval rule, Borda rule etc), allowing abstention or allowing for supermajority rules will not alter our characterization in any way. The other extension is that show that our results can also include the case where voters have general, diverse preferences in addition to diverse information: all we need to do is to change how we interpret the primitives in order to accommodate variation in preferences in our analysis of feasibility of FIE. 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. Hence, while our negative results about continuous state spaces go through, our positive results do not.

Our work helps extend and better understand the existing body of work on information

hyperplanes intersect, they do so inside the simplex.

aggregation in several different ways. We have already discussed how our paper is related to the literature that provide game theoretic proofs of CJT. In addition, our characterization can be used to identify which results are robust to small perturbations of the preference environment. Our results say that the models that use discrete formulation (typically binary state spaces) are robust.[4] On the other hand, the proofs that employ continuous state spaces (e.g., Feddersen-Pesendorfer (1997), McMurray (2017)) are heavily dependent on the particular structure, in particular the assumption of ordered state space.

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)), and so forth. Our work suggests that complexity of the information structure may itself be a barrier to information aggregation.

There are a few papers that identify perverse preference (Acharya 2016, Bhattacharya 2013) or information structures Mandler (2013) are reasons for aggregation failure. Mandler shows that aggregation can break down in a common values model if the same signal indicates opposite states in different situations. Bhattacharya 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 in two-state environments. Moreover, the failure in these papers is a failure of voter co-ordination due to perverse pivotal inference. In our setting, the failure is one of feasibility.

A critique of the existing game theorertic literature is that it involves hyperrationality of voters. In response, non-equilibrium models of voting behavior have been suggested (e.g., Feddersen and Sandroni 2006). We have already mentioned that we avoid that critique by pointing out that the property of information aggregation does not depend on equilibrium inference but on technical feasibility. In particular, in environments which do not allow FIE, aggregation would not be achievable even if voters *commit* to any strategy profile. Therefore, failure does not depend on the particular assumption on voter behavior.

One way look at the question of information aggregation in elections is to observe that while the information that voters have is potentially multidimensional, the action space is limited by the number of alternatives and the voting rule. Given this asymmetry, do we still get the correct outcome? Our work identifies the connection between preference and information for which we can still aggregate information with general state and signal spaces but a binary action space. In fact, we show that the voting rule is irrelevant for the result.

[4] But, at the same time, our results cast doubts on the meaning of FIE in such models, because FIE may well be artificially obtained by the lumping of rich state variables into two representative states.

A related question is whether communication is necessary to produce the correct outcomes. In common value environments, there is a clear incentive to share information. The common value environments where we show that FIE fails are precisely those environments where deliberation is necessary to improve outcomes. On the other hand, in the settings where FIE is achieved through voting, deliberation is not necessary to reach informational efficiency. Coughlan (2000) can be interpreted as saying that deliberation has a role only when there is preference diversity among voters. We, on the other hand, show that it can also have a role under common preferences if information and preferences are sufficiently complex.

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, on the other hand, show that all scoring rules become equivalent in the limit. The implication of this result is that scoring rules matter either in small committees or under diverse preferences. The result that all non-unanimous threshold rules are equivalent is also present in Feddersen and Pesendorfer (1997) and Gerardi and Yariv (2007).

There is also a parallel literature on information aggregation in common value auctions. While Pesendorfer and Swinkels (1997) shows FIE assuming an ordered state space and informative signals, Mihm and Siga (2017) provide a general positive result in discrete environments. Theorem 2 in our paper draws heavily from Mihm and Siga (2017).

2 Model

Consider n players voting over k alternatives in $A = \{a_1, a_2, ..., a_k\}$.[5] 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. In section 6.1, we demonstrate that our results also hold for a larger class of voting rules (e.g., supermajority rules, approval rules, scoring rules etc).

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 general measure 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 measurable function $u : \Theta \times A \to \mathbb{R}$.

Let X be the set of signals. We also allow X to be a general measure space.[6] Given a state θ, each voter privately draws an independent signal $x \in X$ according to a conditional

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

[6] We denote the sigma-algebra in X by $\mathcal{X}$. For a measure space Y, we will use $\Delta(Y)$ to denote the set of all probability measures defined on (the given sigma-algebra of) Y.

probability distribution $P(\,\cdot\,|\theta) \in \Delta(X)$. We will abuse notation and use the same letter P to denote the prior probability on Θ.[7] Given our independence assumption, we informally refer to P as the information structure of the game.

We denote by

$$\mathcal{A}_i = \{\theta \in \Theta : u(\theta, a_i) > u(\theta, a_j) \text{ for all } j \neq i\}$$

the set of states where alternative a_i is strictly preferred to all other alternatives. We assume that $P(\mathcal{A}_i) > 0$ for all $i = 1, 2, ..., k$, that is, every alternative can be preferred ex-ante with positive probability.

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 measurable functions $\sigma_1, \sigma_2, ..., \sigma_k$ with $\sigma_i : X \to [0,1]$ for $i = 1, ..., k$, satisfying $\sum_{i=1}^{k} \sigma_i(x) = 1$ for all $x \in X$. In short, σ is a behavioral strategy mapping X to $\Delta(A)$, where $\sigma_i(x)$ is understood to be the probability of voting for the alternative a_i 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.

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

$$z_i^{\sigma}(\theta) = \int \sigma_i(x) P(dx|\theta).$$

Throughout the paper, we use notation consistent with X and Θ being infinite unless otherwise specified. However, it is always understood that the integral notation is to be replaced by the summation notation when we consider X to be a finite set. Observe that $z_i^{\sigma} : \Theta \to [0,1]$ is a measurable function for each i and σ.

By the SLLN, when every voter uses the strategy σ then, for every θ, the realized proportion of votes for alternative a_i converges $P(\,\cdot\,|\theta)$-a.e. to $z_i^{\sigma}(\theta)$ as $n \to \infty$. Since our focus is on large electorates, we call $z_i^{\sigma}(\cdot)$ the vote share function for alternative a_i. A crucial observation for our analysis is that the vote share function for each alternative is linear in $P(\,\cdot\,|\theta)$.

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,

[7] 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 concepts as primitives. We remark that, as a transition probability, $P(E|\,\cdot\,)$ is a measurable function for every $E \in \mathcal{X}$.

the most preferred alternative is guaranteed to win with an arbitrarily high probability in an ex-ante sense. As a formal standard for information aggregation, we adapt the idea of Full Information Equivalence defined by Feddersen and Pesendorfer (1997).

Given an environment $\{u, A, \Theta, X, P\}$, we first define W_n^σ, the probability of an error (overall ex-ante likelihood of the most preferred alternative not being elected) induced by a strategy σ in a given game with n players. If along a sequence of games as n increases without bound, keeping the environment fixed, the quantity W_n^σ converges to zero, we say that the strategy σ achieves Full Information Equivalence (FIE) and that the environment allows FIE.

For any strategy profile σ and electorate size 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. Denote the random vector representing the proportion of votes for each alternative as $y = (y_1/n,\ y_2/n, ..., y_k/n)$ where $y_i \in \{1, ..., n\}$ is a random variable representing the number of votes for alternative a_i, $i = 1, 2, ..., k$ and $\sum_{i=1}^k y_i = n$.

Given a strategy profile σ and a profile $x^1, ..., x^n$ of signals, the probability of a vector of vote proportions y is given by

$$p_n^\sigma(y|x^1, ..., x^n) = \sum_{\mathcal{B}(y)} \prod_{\ell=1}^{k} \prod_{m_\ell \in B_\ell} \sigma_\ell(x^{m_\ell}),$$

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

$$p_n^\sigma(y|\theta) = \int p_n^\sigma(y|x^1, ..., x^n) \otimes_{m=1}^n P(dx^m|\theta).$$

Let L_n^i denote the set of values of the vector y where the ith coordinate of y is *not* the unique highest, i.e., alternative a_i is not the sole winner. A wrong outcome is obtained if, in a state where a_i 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

$$W_n^\sigma = \sum_{i=1}^{k} \int_{\mathcal{A}_i} p_n^\sigma(L_n^i|\theta) P(d\theta).$$

We say that informaton is fully aggregated if $W_n^\sigma \to 0$ as $n \to \infty$. That is, 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.

Next, we provide an equivalent definition of FIE which is simpler and more relevant to our analytical framework. Recall that, for a given σ, we can define the expected vote share function $z_i^\sigma(\theta)$ for each alternative a_i. Now, let us define the set of states where alternative a_i is elected almost surely by

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

We then have:

Lemma 1 *A strategy σ achieves FIE if and only if*

$$P(\mathcal{A}_i \setminus \mathcal{A}_i^\sigma) = 0.$$

for $i = 1, ..., k$.

That is, σ achieves Full Information Equivalence (FIE) if the set of states where the preferred alternative fails to win almost surely is of prior probability zero. Putting differently, σ achieves FIE if for P-a.e. $\theta \in A_i$, $z_i^\sigma(\theta) > z_j^\sigma(\theta)$ for all $i = 1, ..., k$ and $j \neq i$.[8]

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. Accordingly, if there exists some strategy that aggregates information in a given environment, we will say that the environment *allows* FIE.

We now move to studying environments that allow FIE. As our characterization result will rely on hyperplanes, it will be convenient to work in the space of integrable functions with respect to a given probability measure. So we shall assume the existence of a probability measure $\lambda \in \Delta(X)$ such that $P(\,\cdot\,|\theta)$ is absolutely continuous with respect to λ for P-a.e. θ. Hence, the density $f(\,\cdot\,|\theta)$ of $P(\,\cdot\,|\theta)$ with respect to λ belongs to the Banach space $L_1(\lambda)$ of (equivalence classes of) integrable functions, with the norm $\|f\| = \int |f(x)|\lambda(dx)$. Let $L_1^\Delta(\lambda) = \{f : X \to \mathbb{R}_+ : \int f(x)\lambda(dx) = 1\}$ denote the "simplex" of integrable densities. In the case that X is the finite set $\{x_1, ..., x_s\}$, we take λ to be the uniform distribution $(1/s, ..., 1/s)$, and observe that $L_1^\Delta(\lambda)$ can indeed be identified with the $s-1$ simplex $\Delta(X) = \{(y_1, ..., y_s) \in \mathbb{R}_+^s : \sum_\ell y_\ell = 1\}$ via $(y_1, ..., y_s) \leftrightarrow (\frac{1}{s}y_1, ..., \frac{1}{s}y_s)$. We shall sometimes abuse terminology and refer to both $\Delta(X)$ and $L_1^\Delta(\lambda)$ simply as the simplex.

[8]By P-a.e. $\theta \in \mathcal{A}_i$ we mean for all $\theta \in \mathcal{A}_i$ except for a set of P-measure zero.

Our geometric ideas will belong to the simplex. In particular, we shall be interested in separating the sets $\mathcal{A}_i^\Delta$, $i = 1, ..., k$, given by

$$\mathcal{A}_i^\Delta = \{f(\,\cdot\,|\theta) \in L_1^\Delta(\lambda) : \theta \in \mathcal{A}_i\},$$

and we repeat once more that $\mathcal{A}_i^\Delta$ is simply $\{P(\,\cdot\,|\theta) \in \Delta(X) : \theta \in \mathcal{A}_i\}$ when X is finite. We use the Borel sigma-algebra in $L_1(\lambda)$ and assume that the mapping $\theta \mapsto f(\,\cdot\,|\theta)$ is measurable both ways. As such, the sets $\mathcal{A}_i^\Delta$ and $\{\theta : f(\,\cdot\,|\theta) \in E\}$, for E measurable in $L_1(\lambda)$, are themselves measurable. For such separation ideas, we will use hyperplanes in $L_1(\lambda)$ and their corresponding restrictions to $L_1^\Delta(\lambda)$. A hyperplane in $L_1(\lambda)$ is given by the set $\{g \in L_1(\lambda) : \int g(x)h(x)\lambda(dx) = c\}$, for some $c \in \mathbb{R}$ and bounded measurable $h : X \to \mathbb{R}$, known as the normal of the hyperplane. Our objects of interest are the restrictions of hyperplanes to the simplex,

$$H(h) = \{g \in L_1^\Delta(\lambda) : \int g(x)h(x)\lambda(dx) = 0\},$$

where we note that we gain one degree of freedom, so it is without loss to take c to be zero.[9] Associated with $H(h)$, we consider the positive half-spaces $H^+(h) = \{g \in L_1^\Delta(\lambda) : \int g(x)h(x)\lambda(dx) \geq 0\}$ and $\mathring{H}^+(h) = \{g \in L_1^\Delta(\lambda) : \int g(x)h(x)\lambda(dx) > 0\}$, with the negative ones, $H^-(h)$ and $\mathring{H}^-(h)$, defined analogously. We remark that, because $h \in L_\infty(\lambda)$, the hyperplane with normal h is closed in $L_1(\lambda)$ (and so is its restriction to $L_1^\Delta(\lambda)$, $H(h)$).

3 Feasibility of FIE: Up to Three Alternatives

We start with the case of up to three alternatives, i.e., $k \leq 3$. We shall be able to provide a sharp characterization of FIE. Such result does not apply for the case with $k > 3$, so we study it separately in the next section. In addition to the characterization, $k \leq 3$ seems to be the most relevant case. For instance, the binary case ($k = 2$) is, aside from empirical relevance, the case studied by almost the entire literature on CJT, so results for the binary case are useful in comparison with the rest of the literature.

To demonstrate the condition that determines whether an environment allows FIE or not, we start with two examples in the binary $k = 2$ case. In both these examples we show how FIE may fail when we depart from the two-state framework: the problem arising from the fact that the same alternative can be optimal under multiple conditional distributions over signals. Example 1 shows how the addition of a third state may make FIE infeasible, and Example 2 extends this idea to continuous states.

Given a strategy $\sigma = (\sigma_1, \sigma_2)$, we write the difference in expected vote share between a_1

[9] Simply use $\tilde{h}(x) = h(x) - c$ in the place of h.

and a_2 as

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

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

In order for FIE to obtain, the function $z^\sigma_{12}(\theta)$ 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 $p > \frac{1}{2}$,*

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

This is illustrated in Figure 1 below.

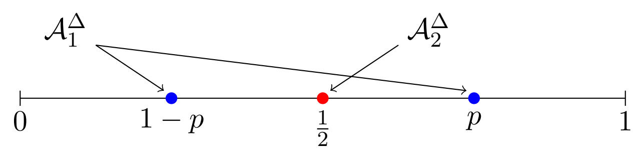


Figure 1

This environment does not allow FIE. In fact, in order for a_1 to win almost surely in both L and R, the strategy σ must satisfy

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

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

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

The problem with information aggregation in this example is the following: In order for a_1 to win in state L (when x is the more frequent signal), voters should vote for a_1 with large enough probability if the signal is x. Similarly, in order for a_1 to win in state R (when y is the more frequent signal), voters should vote for a_1 with sufficiently high probability if the signal is y. As a consequence, a_1 obtains a high share of votes irrespective of the signal, and wins in state M where it is not the preferred alternative.

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 environment, 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* $\theta > t_2$ *These are illustrated in Figure 2 below as cases (A) and (B) respectively. Case (A) allows FIE but case (B) does not. In fact, for any strategy* $\sigma = (\sigma_1, \sigma_2)$, *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.

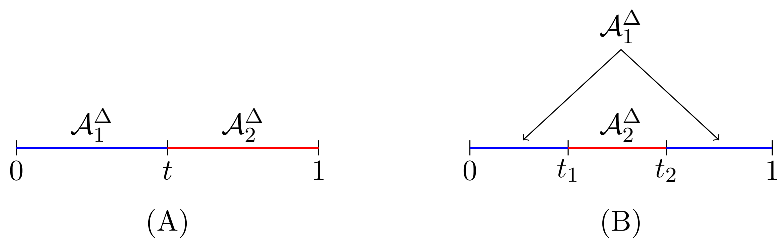


Figure 2

For σ *to satisfy FIE in case (A), we must have* $z_{12}^{\sigma}(\theta) > 0$ *for* $\theta > t$ *and* $z_1^{\sigma}(\theta) < 0$ *for* $\theta < t$. *Hence, any* σ *that satisfies (i)* $z_{12}^{\sigma}(t) = 0$ *and (ii)* $z_{12}^{\sigma}(\cdot)$ *is strictly increasing leads to FIE. It is easy to check that we can always find some* σ *with these properties.*

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

Example 2 says that the substantive interpretation of the signal matters for the property of information aggregation. Suppose voter preferences depend on candidate quality and higher proportion of x (y) signals indicate higher (lower) relative quality of candidate a_1, which is an interpretation of the first environment. This environment allows FIE. However, if voter preferences depend on whether the candidate is moderate or extreme while signals are about whether a candidate leans to the left or to the right, signals cannot be classified as each favoring one candidate. This interpretation applies to the second environment, and aggregation fails in this case.

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 $\mathcal{A}_1^{\Delta}$ and $\mathcal{A}_2^{\Delta}$ are convex, while in the second environment the set $\mathcal{A}_2^{\Delta}$ is non-convex. As we shall see below, convexity of the sets $\mathcal{A}_i^{\Delta}$ is the key feature of an environment that allows FIE.

A *convex 3-partition* of $\Delta(X)$ is given by a collection $\pi = \{E_1, E_2, E_3\}$ of mutually disjoint and convex subsets of $\Delta(X)$ such that $E_1 \cup E_2 \cup E_3 = \Delta(X)$. Given π, we define the *ij-meet* $H_{ij}(\pi) = \overline{E}_i \cap \overline{E}_j$, where $\overline{E}$ denotes the closure of E in the simplex (recall once more that the simplex is either the finite-dimensional simplex $\Delta(X)$ with its canonical Euclidean norm and metric, or the closed subspace $L_1^\Delta(\lambda)$ of the Banach space $L_1(\lambda)$, also with its usual norm and associated metric). In addition, we define the *ij-facet* $F_{ij}(\pi)$ to be equal to $M_{ij}(\pi)$ if there exists $g \in M_{ij}(\pi)$ and $\varepsilon > 0$ such that $B_\varepsilon(g) \subset E_i \cup E_j$, and to be equal to the empty set otherwise.[10] Finally, we define the *ij-facet hyperplane* as $H_{ij}(\pi) = H(h)$ where $H(h)$ is a hyperplane that contains the ij-facet $F_{ij}(\pi)$, whenever such facet is not empty (if $F_{ij}(\pi) = \emptyset$, we set $H_{ij}(\pi) = \emptyset$ as well).

Definition 1 *A convex 3-partition $\pi = \{E_1, E_2, E_3\}$ is called a* **restricted 3-partition** *if for all $i, j, m \in \{1, 2, 3\}$, $H_{ij}(\pi) \neq H_{im}(\pi)$.*

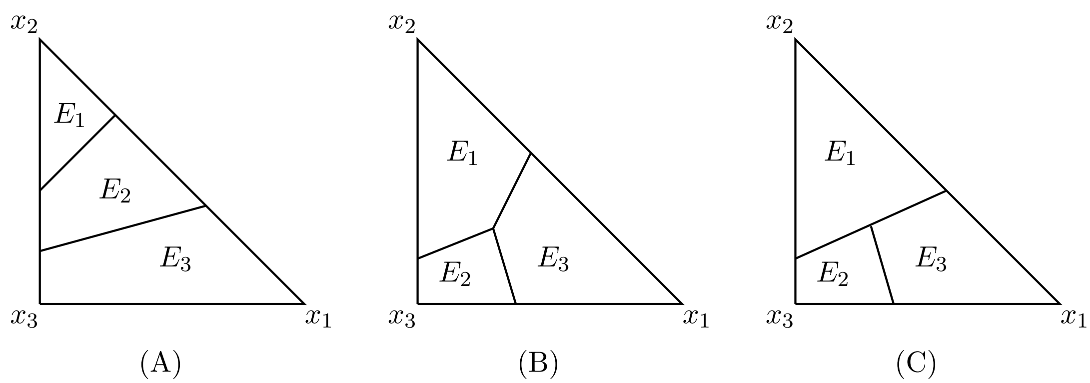


Figure 3

Figure 3 illustrates the concept of a restricted 3-partition for the case that $X = \{x_1, x_2, x_3\}$. The cases (A) and (B) represent valid partitions, while (C) is not a valid partition: the facet hyperplanes H_{12} and H_{13} are equal.

For a given $E \subset \Delta(X)$, let $\mathring{E}$ denote its relative interior in $\Delta(X)$. Here's our characterization result.

Theorem 1 *Let $k = 3$. An environment (u, A, Θ, X, P) allows FIE if and only if there exists a restricted 3-partition $\pi = \{E_1, E_2, E_3\}$ such that*

$$P(\{\theta \in \Theta : f(\cdot|\theta) \in \mathcal{A}_i^\Delta \setminus \mathring{E}_i\}) = 0$$

for $i = 1, 2, 3$.

In simple words, FIE is possible if and only if the images of the three sets $\mathcal{A}_1$ and $\mathcal{A}_2$ and $\mathcal{A}_3$ in the simplex are included in a convex partition defined by three distinct hyperplanes,

[10] $B_e(g)$ is the open ball with radius $\varepsilon > 0$ around g.

with probability one. The case $k = 2$ is much simpler: the required partition into two convex sets is obtained by a single hyperplane, so the characterization boils down to the following corollary.

Corollary 1 *Let $k = 2$. An environment (u, A, Θ, X, P) allows FIE if and only if there exists a hyperplane $H(h)$ such that*

$$P(\{\theta \in \Theta : f(\cdot|\theta) \in \mathcal{A}_1^\Delta \setminus \mathring{H}^+(h)\}) = 0 = P(\{\theta \in \Theta : f(\cdot|\theta) \in \mathcal{A}_2^\Delta \setminus \mathring{H}^-(h)\}).$$

Recalling that the vote share function is linear in $P(\cdot|\theta)$, the intuition for Theorem 1 and Corollary 1 follows on the lines of Example 2. For instance, consider $k = 3$ and a restricted 3-partition as in Figure 3(A). The line separating E_1 and E_2 is a hyperplane with normal h_{12} and the one separating E_2 and E_3 is a hyperplane with normal h_{23}. A strategy $(\sigma_1, \sigma_2, \sigma_3)$ with $\sigma_1 - \sigma_2$ proportional to h_{12} and $\sigma_2 - \sigma_3$ proportional to h_{23} is easily shown to achieve FIE. If the partition is as in Figure 3(B), then one of the normals of the three hyperplanes will be a convex combination of the other two. That is, $h_{13} = \alpha h_{12} + (1 - \alpha)h_{23}$ for some $\alpha \in (0, 1)$. One can then verify that a strategy $(\sigma_1, \sigma_2, \sigma_3)$ with $\sigma_1 - \sigma_2$ proportional to αh_{12} and $\sigma_2 - \sigma_3$ proportional to $(1 - \alpha)h_{23}$ achieves FIE. Conversely, if a profile σ achieves FIE, then by linearity of the vote share function, the differences $\sigma_1 - \sigma_2$, $\sigma_1 - \sigma_3$, and $\sigma_2 - \sigma_3$ must induce a convex partition of $\Delta(X)$. And it cannot be as in Figure 3(C) because otherwise two of the alternatives would always get the same share, which would violate FIE, so it has to be a restricted 3-partition.

For $k = 2$, the ideas are exactly analogous. In particular, FIE is achievable if and only if the images of $\mathcal{A}_1$ and $\mathcal{A}_2$ in the simplex are included in a division of the simplex into two parts by a hyperplane, with probability one. 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 (resp. lower) half-space 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. Figure 4(A) illustrates an environment with discrete states that allows FIE. The blue dots are distributions where, say, a_1 is preferred, while the red dots are distributions where a_2 is preferred. The line dividing the blue and red distributions 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, the higher the probability of signals x_1 and x_3, the more votes a_2 receives. 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 4(B) shows an example where all distributions over signals may arise. In this case there is a unique separation of signals congruent with FIE. Figure 4(C) illustrates a case where the blue and red distributions cannot be separated by a hyperplane, and hence a case where FIE fails.

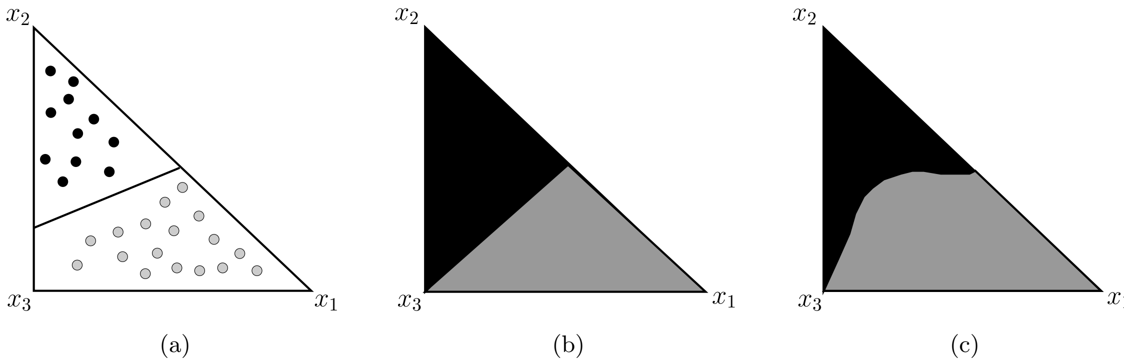


Figure 4: Illustrating Corollary 1.

Corollary 1 identifies an important connection between unidimensional and multidimensional models so far as the property of FIE is concerned. Any multidimensional environment satisfying FIE is characterized by a direction in the simplex such that the projections of the conditional distributions along that direction induce a structure like the first environment in Example 2: there is a threshold that separates the projection of $\mathcal{A}_1^\Delta$ from that of $\mathcal{A}_2^\Delta$. This direction is that of the normal to the hyperplane identified in Corollary 1. Notice that this condition is much weaker than the standard condition of ranking of signals by MLRP.

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 $u_{12}(\theta) = u(\theta, a_1) - u(\theta, a_2)$. Suppose that signal x_1 is optimistic about a_1 and pessimistic about a_2, and vice-versa for the signal x_2. It is reasonable then that $u_{12}(\theta) > 0$ for θ such that $P(\cdot|\theta) = \delta_{x_1}$ and that $u_{12}(\theta) < 0$ for θ such that $P(\cdot|\theta) = \delta_{x_2}$, where δ_x is the point mass at x. It is also reasonable that along a path of indifference (the image of $\{\theta : u_{12}(\theta) = 0\}$ in $\Delta(X)$) the probabilities of x_1 and x_2 should go in opposite directions. An implication of Corollary 1 is that the image of a path of indifference in $\Delta(X)$ is a straight line. In other words, the only utility functions that allow FIE must have linear images in the proportion of each signal.*

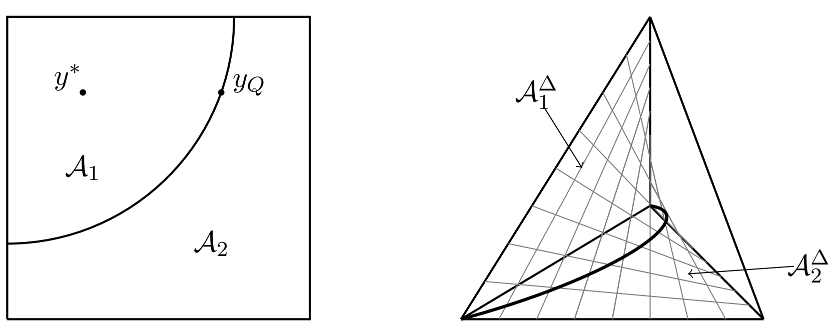
To fix ideas contained in the above example, consider an electorate voting to accept or reject a proposal and that each voter gets noisy information about one of several aspects of the proposal. For concreteness, suppose that the vote is about whether to remain in a common politico-economic union or not (e.g., the "Brexit" vote in May 2016). A crucial tradeoff in this vote was between the possible loss of growth versus protection of local employment. Suppose that 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. Clearly, voter preferences depend on the proportion of each of the three signals in the population. Corollary 1 says that in order for FIE to obtain, μ_3 (the proportion of signals about net contribution of migrants) 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 Corollary 1 to a very standard case of spatial model of political competition between two alternatives. In fact, this is a multidimensional generalization of Example 2.

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. Voters' utility for policy y is given by $u(|y - y^*|)$, $u' < 0$, where $|\cdot|$ denotes the Euclidean norm. Thus, $y^* \in Y$ is the voters' 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 $|y_Q - y^*|$. Hence the boundary between $\mathcal{A}_1$ and $\mathcal{A}_2$ in Θ is given by*

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

which is the circumference of a circle (or part thereof). This is illustrated in Figure 5(A).



(A) (B)

Figure 5

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 $P(x_i = 1|\theta) = \theta_i$. Thus, x_1 provides information on θ_1 and x_2 on θ_2 independently of each other. Alternatively, the ith component of the state, θ_i, can simply be thought of as the proportion of 1-signals in dimension i. Observe that we have four possible combinations of signals, so $\Delta(X)$ is the 3-dimensional simplex, as in Figure 5(B). There is no strategy profile for which FIE can be obtained in this setting. This follows by noting that there exists no hyperplane that can separate $\mathcal{A}_1^\Delta$ from $\mathcal{A}_2^\Delta$ and then applying Corollary 1. Figure 5(B) illustrates this last point. The range of the mapping from Θ to the simplex is the two-dimensional manifold in gray. The boundary $\mathcal{I}$ is mapped to the thick curve dividing the said manifold into $\mathcal{A}_1^\Delta$ and $\mathcal{A}_2^\Delta$. As the image on $\mathcal{I}$ in $\Delta(X)$ is an one-dimensional manifold, it could still be the case that it is contained in a two dimensional hyperplane in $\Delta(X)$. But simple manipulations show that this is not the case.[11]

Both Examples 3 and 4 feature continuous state spaces but finite number of signals. We shall refer to such environments as having "rich state spaces". With rich state spaces, voters have very limited private information but their preferences have a rich variation across different circumstances (states), leading to aggregation failure despite common preferences. FIE holds only for special classes of preference when the state space is rich.

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 3). 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 that 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 $\{\theta_1, \theta_2, ..., \theta_r\}$ occurring with positive probability and s signals $\{x_1, ..., x_s\}$. Assume that the ranking over the two alternatives is strict at each state, and that each state is mapped to a distinct probability distribution $P(\cdot|\theta_t)$ denoted as P_t.

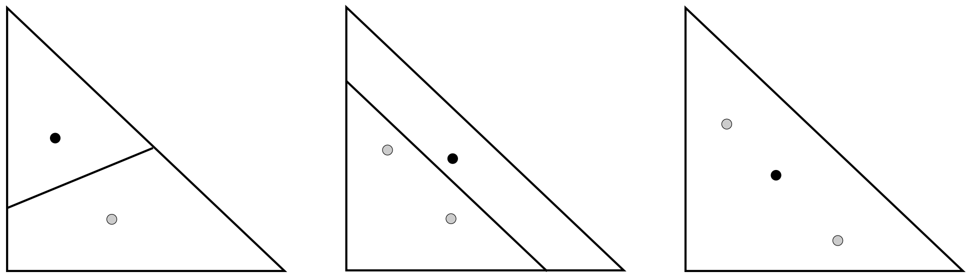
Example 5 *FIE obtains whenever $r = 2$ and $P_1 \neq P_2$.*

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

We skip the proofs as these are special cases of Corollary 3 introduced later. However, for these cases, the statements can simply be verified by inspecting the figures. Figure 6(A)

[11] For instance, in the case that $\mathcal{A}_1 = \{\theta \in [0,1]^2 : \theta_2 > \theta_1^2\}$, the image of $\mathcal{I}$ contains the points $(0,0,0)$, $(1,0,0)$, $(1/8, 3/8, 1/8)$, and $(1/27, 8/27, 1/27)$, and there's no hyperplane in $\mathbb{R}^3$ containing all these points.

illustrates the case of $r = 2$ and $s = 3$: 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 the case of two states and two signals.



(a) (b) (c)

Figure 6: Discrete state space.

Similarly, Figure 6(B) illustrates the case of $r = s = 3$ when the conditional distributions 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 6(C) 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. Notice that in Example 1 also, we have aggregation failure due to the fact that the conditional distribution for which a_2 is preferred is a linear combination of the conditionals for which a_1 is preferred.

4 Feasibility of FIE: Multiple Alternatives

The characterization for $k \leq 3$ does not work for $k > 3$. That is, as Example 7 shows, already with four alternatives a nice convex partition of $\Delta(X)$ made up of the sets $\mathcal{A}_i^{\Delta}$, $i = 1, ..., 4$ is not enough to ensure that FIE is achievable.

Example 7 *Suppose there are three signals and four alternatives. The simplex over the signals is represented by the right angled triangle* ABC*, as shown in Figure 7. 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 vectors in the simplex 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* $\mathcal{A}_i^{\Delta}$ *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* $P(\cdot|\theta)$ *on* AD *(and the entire line along* AD *on the simplex)* $z_1^{\sigma}(\theta) = z_2^{\sigma}(\theta)$. *Similarly, for all points on the line along* BE, $z_3^{\sigma}(\theta) = z_1^{\sigma}(\theta)$. *By linearity of vote shares in* $P(\cdot|\theta)$ *if* BE *and* AD *intersect at* H *then at* $P(\cdot|\theta) = H$, $z_1^{\sigma}(\theta) = z_2^{\sigma}(\theta) = z_3^{\sigma}(\theta)$ *Similarly,*

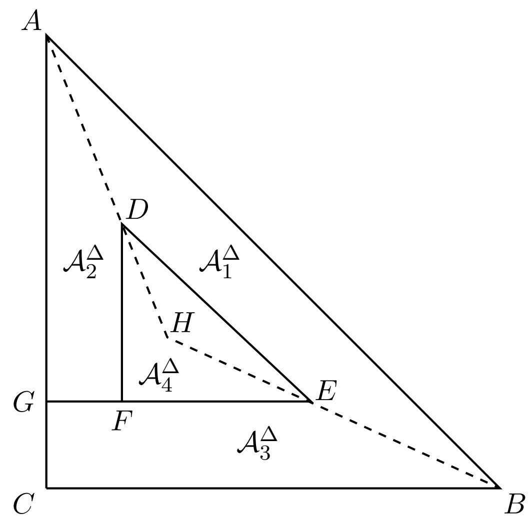


Figure 7: Convex Partition

at F which is the intersection of DF and EF, we must have $z_4^\sigma(\theta) = z_2^\sigma(\theta) = z_3^\sigma(\theta)$. Again 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.

Of course, by linearity of the vote share function, if FIE is to be obtained, then the sets $\{\mathcal{A}_i^\Delta\}_{i=1,...,k}$ must be contained in a convex partition of the simplex. Example 7 shows that convexity is not sufficient for FIE. There are two related issues at stake. First, consider that we have a convex partition of the simplex so that for each pair of alternatives (i,j) the sets $\mathcal{A}_i^\Delta$ and $\mathcal{A}_j^\Delta$ can be separated by a hyperplane $H(h_{ij})$. If a strategy σ achieves FIE, then $\sigma_{ij} = \sigma_i - \sigma_j$ must be proportional to h_{ij}, for each pair (i,j). But we have ${}_kC_2$ hyperplanes $H(h_{ij})$ and only k functions $\sigma_1, ..., \sigma_k$. That is, k unknowns to satisfy ${}_kC_2$ restrictions, and this is not necessarily possible unless we impose extra conditions. Second, the normals h_{ij} need not be linearly independent. This is not an issue with $k = 3$, because in fact we can explore linear dependence among three normals. But for $k > 3$, linear dependence might preclude FIE, as in Example 7.

We shall provide two alternative sets of sufficient conditions for FIE when $k > 3$. Informally speaking, our conditions will be generalizations for the situations described in Figure 3(A) and Figure 3(B). For the former, we will require that the simplex be partitioned in a very nice convex way, so nice that the required hyperplanes are to be parallel. For the latter, we will allow for a more general convex partition, but one that avoids situations like the one in Example 7. In particular, we will require that *for each pair of alternatives* a_i and a_j, the set of conditionals for which a_i is preferred to a_j are separated by a hyperplane from those for which a_j is preferred to a_i.[12]

The reason for this two-step approach is that our first condition, albeit restrictive at first sight, is guaranteed whenever the conditional distributions are linearly independent in the simplex. And, with more signals than states, linear independence holds generically. Of course, linear independence cannot be expected with more states than signals, a result that has a counterpart in the general case. This brings us to our second, more general, condition.

Comparing the two sets of results, we derive the broad lesson that the property of FIE depends on a comparison of the richness of agents' private information with the richness of the underlying preference. This lesson is already reflected in the examples presented in the section with two alternatives: FIE is non-generic when the state space is rich (Examples 3 and 4) and generic when there are at least as many signals as states (examples 5 and 6). In this sense, one may say that there is a positive result for discrete state spaces and a negative result for continuous state spaces.

4.1 Separation by Parallel Hyperplanes

Our first take on the $k > 3$ case is to consider extremely well-behaved convex partitions of the simplex. This is captured by Property PS, which requires that the sets $\mathcal{A}_i^\Delta$, $i = 1, ..., k$, be separated by a set of *parallel* hyperplanes, similarly to the case depicted in Figure 3(A). While this condition seems demanding, under some regularity conditions it is satisfied whenever the conditionals are linearly independent as we shall shortly see.

Definition 2 *We say that* ***property PS*** *holds if there exists a bounded measurable function* $h : X \to \mathbb{R}$ *and real numbers* $0 < c_0 < c_1 < c_2 < \cdots < c_k$ *such that for all* $i = 1, ..., k$, *and* P*-a.e.* $\theta \in \mathcal{A}_i$, $\frac{1}{c_i} < \int h(x) f(x|\theta) \lambda(dx) < \frac{1}{c_{i-1}}$.

Following on the footsteps of Theorem 3 in Siga and Mihm (2017), we have the following

Theorem 2 *If property* ***PS*** *holds, then there exists a strategy that achieves FIE.*

The intuition for the proof is the following. Property PS ensures that for each pair of alternatives (i, j), the sets $\mathcal{A}_i^\Delta$ and $\mathcal{A}_j^\Delta$ are separated by (a translate of) a hyperplane H_{ij} with a common normal h. 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.

While the property may look very demanding, it is satisfied when the set of conditionals satisfy a notion of vector independence. We first present the regularity conditions required for this result to go through.

[12] Observe that this is not true in Example 7, as the separation of alternatives a_2 and a_3 has to have a kink at F: it has to go through G, F, and H.

Definition 3 *We say than an environment* (u, A, Θ, X, P) *is* ***regular*** *if: (i)* Θ *and* X *are compact metric spaces endowed with their Borel sigma-algebras; (ii) the density* $f(\cdot|\theta)$ *of* $P(\cdot|\theta)$ *with respect to* λ *is continuous on* $X \times \Theta$.

Let us denote by $\mathcal{F}$ the set of information structures defined over a regular environment. Observe that when the state and signal spaces are finite, these assumptions are trivially satisfied, so a discrete environment is regular. Formally, letting $\mathcal{M}(\Theta)$ denote the set of all $[-1, 1]$-valued signed measures defined on the Borel sets of Θ, we use the following notion of independence. This notion is a strengthening of the notion of independence in McAfee and Reny (1992); their notion is akin to convex independence, whereas the following notion is akin to linear independence.

Definition 4 *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 signed measure $\nu \in \mathcal{M}(\Theta)$ *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 to 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 following Lemma says that as long as the environment is regular, all we have to verify for property PS is whether the conditionals satisfy the above notion of independence.[13]

Lemma 2 *In a regular environment, if* $P \in \mathcal{F}$ *satisfies independence then property* ***PS*** *holds.*

We thus have the following corollary establishing that linear independence is sufficient for a regular environment to allow FIE. Notice that independence is a property of the information structure: as long as the conditionals satisfies independence, FIE obtains irrespective of preferences.

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

[13] 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 *linearly dependent* 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.

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

Corollary 3 *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 linearly independent.*

The above Corollary generalizes Examples 5 and 6. Of course, when $s \geq r$ linear independence is a generic property. The general lesson with finite state and signal spaces and *with more signals than states* is that FIE obtains except for very special cases. We postpone a formal discussion of genericity of FIE till section 4.1.1.

Corollary 3 is important given the large body of work looking specifically at the case with two states ($r = 2$). It is easy to see that when there are two states, linear independence is trivially satisfied as long as $P_1 \neq P_2$, i.e., each state produces a different conditional distribution over signals.

With discrete environments with more states than signals ($r > s$), linear independence fails and Corollary 2 cannot offer any guidance regarding whether FIE holds or not. However, there is a sense in which more states and/or fewer signals is an impediment for FIE. The point is demonstrated by the following heuristic argument due to Siga and Mihm (2017) Consider $r > s$ and $k = 2$. Then, the condition for FIE is given by Theorem 1. Now, suppose we increase the number of states r keeping the number of signals s fixed. For each additional state, we choose a random vector on the simplex as the relevant conditional distribution and assign a random alternative as the most preferred one. As states are added to the simplex in this manner, the (ex-ante) likelihood of the condition in Theorem 1 being violated increases. In fact, one can make the likelihood of FIE obtaining arbitrarily small by sufficiently increasing the number of sates (see Mihm and Siga (2017), Theorem 3, for a formal statement).

The logical limit of the procedure above is to consider situations with infinitely many states and finitely many signals. For instance, by having a "rich state space" with the entire $s-1$ simplex as the range of the mapping $\theta \mapsto P(\,\cdot\,|\theta)$. In such a situation, linear independence cannot be expected, so in Section 4.2 below we develop a weaker set of sufficient conditions for FIE which apply to rich state spaces. But let us first delve into the idea of genericity for general spaces Θ and X.

4.1.1 Genericity of FIE

In general environments, state and signals spaces are bound to not be finite. So generic independence (and the resulting generic FIE property) when there are more signals than states cannot be taken as an indication of prevalence of FIE. We now argue that *failure of FIE* might well be robust.

We demonstrate the failure of genericity of FIE with two examples. Both examples involve two alternatives in order to draw from the characterization in Corollary 1. The first example shows that even in a discrete environment involving more states than signals, FIE can fail for an open set of environments.

Example 8 *Let $X = \{x_1, x_2, x_3\}$ and $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$. Let there be two alternatives, with $\mathcal{A}_1 = \{\theta_4\}$ and $\mathcal{A}_2 = \{\theta_1, \theta_2, \theta_3\}$. Consider an information structure $P : \Theta \to \Delta(X)$ such that $P_i = e_i$ for $i = 1, 2, 3$, and $P_4 = (1/3, 1/3, 1/3)$, where e_i is the coordinate vector (i.e. $e_1 = (1, 0, 0)$, etc.) and P_i is short for $P(\cdot|\theta_i)$. Hence, under P, $\mathcal{A}_1^\Delta$ is the mid-point of the simplex, and $\mathcal{A}_2^\Delta$ is the union of the three vertices, as illustrated in Figure 8(A). There's no way to separate these two sets with a single hyperplane, so FIE fails for the information structure P. Now consider information structures close-by, which here means that $\hat{P}_i$ is close to P_i for each i as vectors in $\mathbb{R}^3$ (restricted to the simplex, of course). It is clear that we can find an open set of such $\hat{P}$'s such that the corresponding images of $\mathcal{A}_1$ and $\mathcal{A}_2$ in the simplex will be close to the center and the vertices, respectively. This is illustrated in Figure 8(B), with the open balls depicted. Again, for close enough $\hat{P}$'s, it will not be possible to separate the corresponding sets with a single hyperplane. So FIE fails for each such $\hat{P}$.*

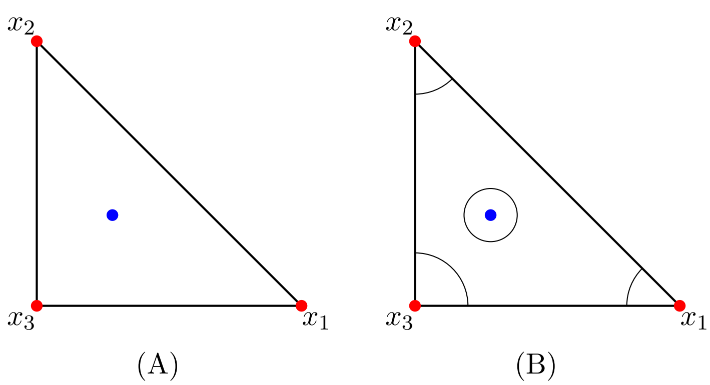


Figure 8

It is straightforward to verify that similar ideas can be applied to Example 2(B), with infinitely many states and two signals. In fact, the ideas in Example 8 already show that FIE can fail robustly for general Θ and X: it suffices to have λ have four atoms at the dots in 8(A). The next example presents an alternative argument for infinite Θ and X.

Example 9 *Let Θ be compact metric, $X = [0, 1]$, and λ be the Lebesgue measure on $[0, 1]$. Let $P : \Theta \to L_1^\Delta(\lambda)$ be an information structures such that $\mathcal{A}_1^\Delta = \{f \in L_1^\Delta(\lambda) : ||f - 1|| < \frac{1}{4}\}$ and $\mathcal{A}_2^\Delta = L_1^\Delta(\lambda) \setminus \mathcal{A}_2^\Delta$. Clearly there's no way to separate $\mathcal{A}_1^\Delta$ and $\mathcal{A}_2^\Delta$ with a hyperplane: for any bounded $h : [0, 1] \to \mathbb{R}$, we would need $\int h(x)dx > 0$ because $1 \in \mathcal{A}_1^\Delta$; hence we would have an interval $I \subset [0, 1]$ with $h(x) > 0$ for all $x \in I$; now take one such interval I such that, for some*

$\theta \in \mathcal{A}_2$, $f(x|\theta) = \frac{1}{\lambda(I)}\mathbb{1}\{x \in I\}$, so that we would have $\int h(x)f(x|\theta)dx > 0$, contradicting separation. Consider the space of all information structures $P : \Theta \to L_1^{\Delta}(\lambda)$ such that the range of P is $L_1^{\Delta}(\lambda)$ and endow it with the metric $\rho(P, \hat{P}) = \sup_{\theta\in\Theta} ||f(\cdot|\theta) - \hat{f}(\cdot|\theta)||$, where $\hat{f}$ is the density of $\hat{P}$. For small $\varepsilon > 0$, pick an ε-ball $B_\varepsilon(P)$ around P, and observe that an analogous argument establishes that, for each $\hat{P} \in B_\varepsilon(P)$, there is no hyperplane separating the corresponding sets $\hat{\mathcal{A}}_1^{\Delta}$ and $\hat{\mathcal{A}}_2^{\Delta}$.

The restriction to surjective information structures (that is, to P with range equal to $L_1^{\Delta}(\lambda)$) in Example 9 is important. Without it, we would not necessarily be able to find an open set of $\hat{P}$'s for which FIE would fail. It can be interpreted as restricting to environments with rich state spaces, something we will assume in Section 4.2 below (in fact, this is exactly what property R will require).

The conclusion from Examples 8 and 9 is that FIE can fail robustly. It is important to stress the kind of independence that is required for property PS. As indicated in footnote 13 above, we need the full force of linear independence to ensure that property PS is satisfied. Applying the analysis of Hellwig and Gizatulina (2017) to our setting shows that the set of information structures satisfying *convex* independence is generic (under some regularity conditions). Example 9 shows that the same is not true for linear independence.

4.2 Separation by Star-Shaped Partitions

We move now to our more general approach to handle the $k > 3$ case. When $\mathring{E}_i$ is equal to $\mathcal{A}_i^{\Delta}$ in the convex partition in Figure 3(B), we see that, for each pair (i, j) of alternatives, the region in the simplex where i is preferred to j is separated by a hyperplane. Moreover, such separating hyperplanes intersect at one single point, forming what one can call a "star-shaped" partition. Of course this is weaker than requiring separation by parallel hyperplanes, and in particular it allows for linearly dependent conditionals.

It turns out that the crucial feature of "star-shapedness" is the common intersection of the associated hyperplanes associated with a triple of alternatives. Such point might not be located in the simplex, though. So we first extend preferences to allow for densities to live outside of the simplex.

Definition 5 *We say that **property E** is satisfied if: (i) there is a measure space $\tilde{\Theta} \supseteq \Theta$ and a signed measure $\tilde{\lambda}$ on the space $\mathcal{M}(X)$ of $[-1, 1]$-valued signed measures on $\mathcal{X}$ such that $\tilde{\lambda}(\tilde{\Theta}) = 1$ and $\tilde{\lambda}|_{\Delta(X)} = \lambda$; (ii) the information structure is described by $\tilde{P}(\cdot|\theta)$ for all $\theta \in \tilde{\Theta}$, which is absolutely continuous with respect to $\tilde{\lambda}$, and coincides with $P(\cdot|\theta)$ for $\theta \in \Theta$ and the prior $\tilde{P}$ on $\tilde{\Theta}$ coincides with P conditional on Θ, i.e., $\tilde{P}(\cdot|\Theta) = P(\cdot)$; (iii) there is a utility function $\tilde{u} : \tilde{\Theta} \times A \to \mathbb{R}$ such that $\tilde{u}|_{\Theta\times A} = u$.*

In the extended environment, we consider the extended simplex $L_1^{\Sigma}(\tilde{\lambda}) = \{f : X \to \mathbb{R} : \int f(x)\tilde{\lambda}(dx) = 1\}$, which reduces to the set Σ of vectors that add up to one in the case of finite X, and the corresponding restriction of hyperplanes: for a given bounded measurable h, $H(h) = \{g \in L_1^{\Sigma}(\tilde{\lambda}) : \int g(x)h(x)\tilde{\lambda}(x) = 0\}$.

Definition 6 *We say that* ***property H*** *is satisfied if for all pair of alternatives $a_i, a_j \in A$, there exists a bounded measurable $h_{ij} : X \to \mathbb{R}$ such that*

$$\tilde{P}(\theta \in \tilde{\Theta} : \tilde{f}(\,\cdot\,|\theta) \in \mathcal{A}_{ij}^{\Sigma} \setminus \mathring{H}^{+}(h_{ij}\}) = 0 = \tilde{P}(\theta \in \tilde{\Theta} : \tilde{f}(\,\cdot\,|\theta) \in \mathcal{A}_{ji}^{\Sigma} \setminus \mathring{H}^{-}(h_{ij}\}),$$

where $\mathcal{A}_{ij} = \{\theta \in \tilde{\Theta} : \tilde{u}(\theta, a_i) > \tilde{u}(\theta, a_j)\}$, $\mathcal{A}_{ij}^{\Sigma}$ is its image on $L_1^{\Sigma}(\tilde{\lambda})$, and $\tilde{f}(\,\cdot\,|\theta)$ is the density of $\tilde{P}(\,\cdot\,|\theta)$ with respect to $\tilde{\lambda}$.

While the necessary condition for FIE (convexity of $\mathcal{A}_j^{\Delta}$) imposes conditions only on the most preferred alternative in each state, property **H** imposes conditions on the entire ranking over alternatives. For any σ, the set of states which produce equal vote shares for a given pair of alternatives is characterized by a hyperplane, irrespective of the vote shares received by the other alternatives for these states. Property **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.

Finally, we shall restrict to the case of rich Θ so as to obtain a sharp set of sufficient conditions.

Definition 7 *An environment satisfies* ***property R*** *if for each $g \in L_1^{\Delta}(\lambda)$ there is $\theta \in \Theta$ such that $f(\,\cdot\,|\theta) = g(\cdot)$.*

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

Lemma 3 *Suppose properties* ***E***, ***H***, *and* ***R*** *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*

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

What Lemma 3 establishes is that, for any three alternatives, the hyperplanes from property **H** must either have a common intersection (which might lie outside of the simplex) or be parallel to each other. Hence property **H** is a substantial weakening of property **PS**.

Figure 9 makes the argument graphically by contradiction. Consider three alternatives $\{a_1, a_2, a_3\}$ and suppose the dashed lines are the hyperplanes ($H(h_{12}), H(h_{13})$, and $H(h_{23})$) from property **H**. Suppose that the result in the Lemma 3 is violated, and the three hyperplanes have three separate pairwise intersections. Note that the colored areas represent the region where an alternative is best. But then for any θ such that the corresponding

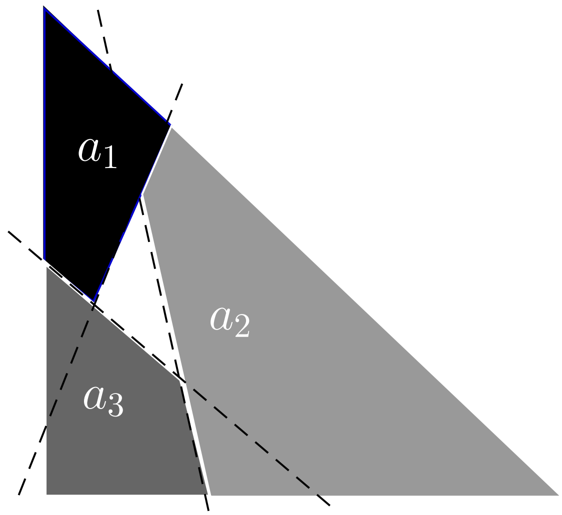


Figure 9: FIE failure with loops.

conditional lies on the inner uncolored triangle features an intransitive preference cycle: $u(\theta, a_1) > u(\theta, a_2) > u(\theta, a_3) > u(\theta, a_1)$.

Now, we are ready to state and prove the main result of this section.

Theorem 3 *If Properties **E**, **H**, and **R** hold, then 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.

4.2.1 Linear utility representation

Our analyzes so far have kept the general structure of a state space Θ, a signal space X and, given an information structure, the implied distribution on the simplex for each state $\theta \in \Theta$. An alternative approach is to view states as the distributions themselves, and define the common utility as a real-valued function defined on the simplex and alternatives. It is apparent that, under this alternative route, FIE is related to the linearity of the utility function in states. We now make this intuition precise.

The set of states is identified as a subset M of $L_1^{\Delta}(\lambda)$ and the utility function is defined directly over M, that is, $u : M \times A \to \mathbb{R}$ is a bounded measurable function, denoted by $u(f, a)$ with $f \in M$ to highlight that states are themselves distributions. We shall denote by $\mathcal{A}(u)$ the partition $\{\mathcal{A}_i^{\Delta}\}_{i=1,\ldots,k}$ of the state space M induced by the utility function u.

Our result is the following. If the utility from each alternative is linear in states in M, then the environment allows FIE. Conversely, for every environment that allows FIE, there exists a utility function linear in states which induces the same top-ranked alternative for each state.

Proposition 1 *If the utility functions is given by $u(f, a) = \int f(x)u_a(x)\lambda(dx)$, where u_a is bounded and measurable for all $a \in A$ and $f \in M$ then the environment allows FIE. Conversely, if a given utility function $u : M \times A \to \mathbb{R}$ belongs to an environment that allows FIE, then there exists another environment with the utility function $\hat{u}(f, a) = \int f(x)\hat{u}_a(x)\lambda(dx)$, where $\hat{u}_a \in L_\infty(\lambda)$, such that $\mathcal{A}(u) = \mathcal{A}(\hat{u})$.*

This characterization is also portrayed in Example 3 and the discussion following it. An interpretation when $M = L_1^\Delta(\lambda)$ (so that property **R** is satisfied) 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.

Proposition 1 holds true for utility functions defined over any subset of the simplex: in particular, it applies to discrete sets M (arising from discrete state spaces) too. However, it has more intuitive value when the signal proportions can be varied continuously. In particular, when property **R** is satisfied, then it can be verified that linearity of the utility function is equivalent to property **E** and **H** being jointly satisfied by the environment. We shall consider rich state spaces for the rest of the discussion on linear utility representation.

We already know that property **E**, **H** and **R** together are sufficient for FIE, which is also reflected in the sufficiency of linearity in Proposition 1. While such properties are not strictly necessary for FIE, the converse in the proposition tells us that for any environment satisfying FIE, there must be another environment with the same top-ranked alternative in each state, satisfying properties **E** and **H** (and hence FIE). It is also worth noting that when $k = 2$, FIE is indeed characterized by linear utility functions.

For environments that admit linear utility representations, we can obtain a classification of signals in terms of which among a given pair of alternatives is favored. Suppose that for any two alternatives a_i and a_j, the respective (linear) utility functions are $u(f, a_i) = \int f(x)u_{a_i}(x)\lambda(dx)$ and $u(f, a_j) = \int f(x)u_{a_j}(x)\lambda(dx)$, respectively. Along the hyperplane $H(h_{ij})$ on the simplex describing indifference between a_i and a_j, we must have

$$\int f(x)[u_{a_i}(x) - u_{a_j}(x)]\lambda(dx) = 0.$$

We can then partition X into $\{X_i, X_j, X_{ij}\}$ by setting $X_i = \{x \in X : u_{a_i}(x) > u_{a_j}(x)\}$, $X_j = \{x \in X : u_{a_j}(x) > u_{a_i}(x)\}$, and $X_{ij} = \{x \in X : u_{a_i}(x) = u_{a_j}(x)\}$. Signals in X_i favor a_i and those in X_j favor a_j in the sense that a higher proportion of any signal in X_i at the expense of any signal in X_j raises the utility difference between a_i and a_j. Moreover, it can be checked that for any strategy σ that achieves FIE, it must be the case that $\sigma_{ij}(x) > 0$ if $x \in X_i$ and $\sigma_{ij}(x) < 0$ if $x \in X_j$.

5 Equilibrium analysis

Summing up, by focusing on the geometry of the conditional distributions over signals, we have established in general conditions under which FIE can and cannot be obtained. But such results deal only with the feasibility of information aggregation, in the sense of existence of some strategy profile that achieves FIE. However it is not clear whether, even in environments which allow FIE, voters have incentives to use such strategies. In order to check whether voters find it in their interest to do so, we consider voting as a game. More precisely, 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. 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. We abuse notation and use the letter a to denote a profile of voter 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.[14] Notice that all voters have the same utility function. 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 strategy 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.

[14] 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.

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 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., $W_n^{\sigma^n} \to 0$.*

The above theorem establishes McLennan's result in our setting: in environments where FIE is feasible, FIE can be achieved by a sequence of Nash equilibria. This result demonstrates that the failure of information aggregation in common value environments is a failure of technical feasibility rather than that of incentive compatibility.

6 Extensions

6.1 Scoring rules

We have developed our conditions for FIE based on simple plurality rule where each voter casts his or her vote for one and only one alternative. However, there are other voting rules to be considered especially when there are more than two alternatives, e.g., approval voting, Borda count, etc. We show that considering these other voting rules does not expand the set of environments where we can aggregate information. In particular, the set of environments for which FIE can be achieved is the same under plurality rule with or without abstention and approval rule. Additionally, whenever FIE is achieved under Borda rule, it is also achieved under plurality rule. Finally, when there are two alternatives, supermajority rules induce FIE if and only if the simple majority rule induces FIE.

Formally, our result is an equivalence result between the plurality rule and a 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. This class includes approval voting as a special case. Other voting rules like the plurality rule (with or without abstention) and Borda rule can be obtained as scoring rules with restrictions on the ballot.

We follow Myerson (2002) for defining a scoring rule. Let X be a finite set of signals and V be a positive integer. A *V-scoring rule* is a voting procedure where a voter can assign any integer score between 0 and V to each alternative. Formally, when there are k alternatives, a voter picks a ballot which is a vector $v \in \mathcal{V} = \{0, ..., V\}^k$, and each element of the ballot, v_j, is interpreted as the score he gives to alternative j. Ballots are aggregated by adding the scores 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. Under the Borda Rule, a voter provides a ranking of the alternatives, and the alternative with the highest aggregate rank wins. One can reinterpret the ranks assigned by a voter as points awarded in the descending order, with the highest ranked alternative obtaining $k-1$ and the lowest ranked one getting 0.

Definition 8 (Approval voting) *An approval voting rule is the scoring rule with* $V=1$.

Definition 9 (Plurality voting) $V=1$ *and a ballot* $v \in \mathcal{V}$ *requires* $\sum_i v_i = 1$.

Definition 10 (Plurality voting with Abstention) $V=1$ *and a ballot* $v \in \mathcal{V}$ *requires* $\sum_i v_i \leq 1$.

Definition 11 (Borda Count) $V=k-1$ *and a ballot* $v \in \mathcal{V}$ *requires that no two alternatives are assigned the same number of points.*

At this point, it is important to distinguish between "pure" scoring rules and scoring rules with balloting restrictions. While approval rule belongs to the former class, plurality rule, plurality with abstention and Borda rule belong to the latter group. Notice that, for any given V, if an environment allows FIE under a V-scoring rule with balloting restrictions, it also allows FIE under the pure V-scoring rule since the strategy achieving FIE under the former rule is also available under the latter.

The next result tells us that scoring rules (with or without restrictions) cannot do more than the plurality rule in terms of delivering FIE: For "pure" scoring rules like the approval rule, FIE is achieved if and only if FIE is achieved under plurality rule, FIE under scoring rules with restrictions (e.g. Borda Rule) implies FIE under plurality rule. The theorem also simultaneously establishes the equivalence of all V-scoring rules as far as the property of FIE is concerned.

Theorem 5 *Fix* $V \in \mathbb{N}_+$. *There exists a strategy that achieves FIE in a* V-scoring rule *without restrictions if and only if there is a strategy profile that achieves FIE in plurality voting. If there exists a strategy that achieves FIE in a* V-scoring rule *with balloting restrictions, then there is a strategy profile that achieves FIE in plurality voting.*

The above result comes with two caveats. First, this 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 2 *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 Monotone Likelihood Ratio Property

In our framework, we obtain conditions on FIE with general signal and state spaces. One way to compare our result to existing work 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 show (albeit in a model of diverse preferences) that information is aggregated in all equilibria. Our sufficient condition for FIE adapted to this environment entertains MLRP as a special case. Let us restrict to the two alternative case, as the extension to multiple alternatives is immediate.

We start by making the following formal assumptions. Suppose $\Theta = [0, 1]$ and $X = \{(x_1, ..., x_k) : (x_1, ..., x_k) \in [0, 1]^k, \text{ with } x_1 < x_2 < \cdots < x_k\}$. The prior P over $[0, 1]$ is non-atomic and has full support. The preferences are as follows: for some $\theta^* \in (0, 1)$, a_1 is preferred for $\theta > \theta^*$ and a_2 is preferred for $\theta < \theta^*$. In this setting, MLRP is defined as the following condition on $P(\cdot|\cdot)$.

Definition 12 (Monotone Likelihood Ratio Property) *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.*

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 $\theta' > \theta^*$ and each $\theta'' < \theta^*$, 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.

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 $(\theta^*, 1)$ or within $(0, \theta^*)$.[15]

6.3 Diverse Preferences

So far we have assumed that all voters have the same preferences described by the common utility function $u(\theta, a)$. In this section we extend our basic insight to a case where the voters in the electorate may have different preferences. In this setting, our results on feasibility go through (almost exactly as before, with a different interpretation of the primitives). However, feasibility of FIE cannot guarantee information aggregation in equilibrium since McLennan's insight fails with diverse preferences.

We maintain the assumption that all voters are ex ante identical, and draw both their information and preferences from some distribution conditional on the state. To do so, we retain the elements of the set-up 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 captured by $u : \Theta \times X \times A \to \mathbb{R}$. Given u and P, we can infer the underlying "common" preference of a large electorate, as follows. First and for simplicity, let us assume that for every pair of alternatives (i, j), almost every θ and $P(\cdot|\theta)$-a.e. x, $u(\theta, x, a_i) \neq u(\theta, x, a_j)$. By the SLLN, the number

$$Q_{ij}(\theta) = \int \mathbb{1}\{u(\theta, x, a_i) > u(\theta, x, a_j)\} P(dx|\theta)$$

represents the proportion of the electorate that prefers a_i to a_j in state θ. Hence, when $Q_{ij}(\theta) > \frac{1}{2}$, alternative a_i would get more than 50% of the votes if the state was known. We advance that in such a state the electorate prefers a_i to a_j. So we set, for each pair (i, j),

$$\mathcal{A}_{ij} = \{\theta \in \Theta : Q_{ij}(\theta) > \frac{1}{2}\},$$

[15] An analogous result is obtained by Mihm and Siga (2017), who show that, in order for information to be aggregated in common-value auctions, information must be monotone with respect to a "betweenness order", which is strictly weaker than the ordering induced by MLRP.

and

$$\mathcal{A}_i = \bigcap_{j \neq i} A_{ij}.$$

Armed with these sets, we can extend the definition of FIE to the requirement that there exists a symmetric profile σ such that $P(\mathcal{A}_i \backslash A_i^\sigma) = 0$ for every $i = 1, ..., k$, as before. Likewise, using the images of the sets $\mathcal{A}_i$ and $\mathcal{A}_{ij}$ in the simplex (or in the extended simplex), we can immediately recast the definitions of *restricted 3-partition* and of *properties* **PS** and **H**, and conclude that Theorem 1, Corollary 1, Theorem 2, and Theorem 3 remain valid in this more general setting.

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.

Observe that 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 focused on specific preference and information environments. We instead consider general environments with arbitrary preference and information structures and focus on properties of the environment allowing or precluding information aggregation. The main thrust of our analysis is the focus on the geometry of the sets of probability distributions over private signals corresponding to the partition of the state space induced by the common state-dependent utility function of the voters. 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 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.

Most of our analysis assumes the existence of a common utility function, so there is no issue of preference aggregation, only of information aggregation. We provide a complete characterization of feasibility of FIE for the case of up to three alternatives. Roughly speaking, the partition of the state space induced by the preferences is to be represented in the simplex of distributions over signals as a "nice" partition into convex polytopes with facets defined by hyperplanes. For the case of more than three alternatives, we do not have such a sharp characterization. Instead, we provide two sets of sufficient conditions. The first requires that the said hyperplanes be parallel and the second allows for more general, "star-shaped", configurations of hyperplanes. Interestingly, the first condition holds generically when we have more signals than states. However, with general state and signal spaces FIE can fail robustly.

We provide an affirmative answer to the implementability issue: as long as an environment allows FIE, there is a sequence of equilibria associated with ever increasing electorates that achieves FIE. There may be other equilibrium sequences that do not aggregate information—but ours is only a possibility result. A corollary is that FIE is always achieved in equilibrium in the much studied two-state environment. Another corollary is that 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: whenever information *can be* aggregated, information *will be* aggregated in (some) equilibrium. 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. On the other hand, although our feasibility results extend to the case of diverse preferences, such extension is not available for our equilibrium result.

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 would have incentives to share their private information. Therefore, information would trivially be aggregated. In this context, our positive results are significant. In particular, if the number of signals is larger than the number of states, then information aggregation does not require communication, in general. 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.

References

[1] Acharya, Avidit. "Information Aggregation Failure in a Model of Social Mobility," Games and Economic Behavior, Volume 100, November 2016, pp. 257-272

[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] Austen-Smith, David and Timothy Feddersen. 2006. Deliberation, Preference Uncertainty and Voting Rules. American Political Science Review. 100(2): 209-217.

[6] Balder, Erik, “Generalized Equilibrium Results for Games with Incomplete Information.” Mathematics of Operations Research 13, (2013): 265-276.

[7] Berg, Sven. “Condorcet’s jury theorem, dependency among jurors.” Social Choice and Welfare 10, no. 1 (1993): 87-95.

[8] Bhattacharya, Sourav. “Preference Monotonicity and Information Aggregation in Elections.” Econometrica 81, no. 3 (2013): 1229-1247

[9] Coughlan, Peter (2000). In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. The American Political Science Review, 94(2):375 393

[10] Duggan, John, and César Martinelli. “A Bayesian model of voting in juries.” Games and Economic Behavior 37, no. 2 (2001): 259-294.

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

[12] Feddersen, Timothy, and Alvaro Sandroni. 2006. ”A Theory of Participation in Elections.” American Economic Review, 96(4): 1271-1282.

[13] Gerardi, Dino and Leeat Yariv. “Deliberative Voting”, Journal of Economic Theory 134, no. 1 (2007): 317-338

[14] Gizatulina, Alia, and Martin Hellwig. “The Generic Possibility of Full Surplus Extraction in Models with Large Type Spaces”Journal of Economic Theory 170, (2017): 385-416.

[15] Ladha, Krishna K. “The Condorcet jury theorem, free speech, and correlated votes.” American Journal of Political Science (1992): 617-634.

[16] Mandler, Michael. “The fragility of information aggregation in large elections.” Games and Economic Behavior 74, no. 1 (2012): 257-268.

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

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

[19] McMurray, Joseph. "Why the Political World is Flat: An Endogenous Left-Right Spectrum in Multidimensional Political Conflict", (2017) mimeo, Brigham Young University

[20] Mihm, Maximillian and Lucas Siga. "Information Aggregation in Large Auctions" mimeo, New York University (2017)

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

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

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

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

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

8 Appendix

8.1 Proof of Lemma 1

To verify that the two definitions are equivalent, say that $P(\mathcal{A}_i \setminus \mathcal{A}_i^\sigma) = 0$ for all i. Then $W_n^\sigma = \sum_i \int_{\mathcal{A}_i^\sigma \cap \mathcal{A}_i} p_n^\sigma(L_n^i|\theta)P(d\theta)$. For each i and $\theta \in \mathcal{A}_j^\sigma \cap \mathcal{A}_i$, we have $z_i^\sigma(\theta) > z_j^\sigma(\theta)$ for every $j \neq i$, and we know that the realized proportion $z_i^n(\theta)$ converges a.s. to $z_i^\sigma(\theta)$. This implies $p_n^\sigma(L_n^i|\theta) \to 0$ for every θ. As this is true for every i, by Lebesgue Dominated Convergence it follows that $W_n^\sigma \to 0$ as $n \to \infty$. Conversely, if $P(\mathcal{A}_i \setminus \mathcal{A}_i^\sigma) > 0$ for some i, then there is a set of positive P-measure $E \subset \mathcal{A}_i$ and an alternative j such that $z_j^\sigma(\theta) > z_i^\sigma(\theta)$ for all $\theta \in E$. Let $P(E) = \alpha > 0$. Then $\liminf_n p_n^\sigma(L_n^i|\theta) \geq \alpha$ because $z_n^\sigma(\theta)$ converges $P(\cdot|\theta)$-a.s. to $z^\sigma(\theta)$. By Fatou's Lemma, W_n^σ cannot converge to 0, so FIE fails.

8.2 Proof of Theorem 1

For the *only if* part. Let σ be a strategy that achieves FIE. Define $E_1 = H^+(\sigma_1 - \sigma_2) \cap H^+(\sigma_1 - \sigma_3)$, $E_2 = \mathring{H}^-(\sigma_1 - \sigma_2) \cap H^+(\sigma_2 - \sigma_3)$, and $E_3 = \mathring{H}^-(\sigma_1 - \sigma_3) \cap \mathring{H}^-(\sigma_2 - \sigma_3)$. We first establish that $\pi = \{E_1, E_2, E_3\}$ is a convex partition, and then establish that it is a restricted 3-partition. First, suppose that $E_1 \cup E_2 \cup E_3 \neq \Delta(X)$, so there is $g \in \Delta(X)$ and not in either of the E's. There are two feasible cases:

- $[g \in \mathring{H}^-(\sigma_1 - \sigma_2) \cap \mathring{H}^-(\sigma_2 - \sigma_3) \cap H^+(\sigma_1 - \sigma_3)]$. But if $g \in \mathring{H}^-(\sigma_1 - \sigma_2) \cap \mathring{H}^-(\sigma_2 - \sigma_3)$ then $g \in \mathring{H}^-(\sigma_1 - \sigma_3)$, contradicting that $g \in H^+(\sigma_1 - \sigma_3)$.
- $[g \in \mathring{H}^-(\sigma_1 - \sigma_3) \cap H^+(\sigma_2 - \sigma_3) \cap H^+(\sigma_1 - \sigma_2)]$. But if $g \in \cap H^+(\sigma_2 - \sigma_3) \cap H^+(\sigma_1 - \sigma_2)$ then $g \in H^+(\sigma_1 - \sigma_3)$, contradicting that $g \in \mathring{H}^-(\sigma_1 - \sigma_3)$.

So we conclude that $E_1 \cup E_2 \cup E_3 = \Delta(X)$. Next, as each E_i is in the complement of one another, they are mutually disjoint. And surely each E_i is convex, so π is a convex partition.

Finally, we show that π must be a restricted 3-partition. By construction, for all E_i, E_j that share a facet, $H_{ij}(\pi) = H(\sigma_i - \sigma_j)$. To show a contradiction and without loss of generality, suppose $H_{12}(\pi) = H(\sigma_1 - \sigma_2) = H(\sigma_1 - \sigma_3) = H_{13}(\pi)$. Now, $H_{12}(\pi) = H_{13}(\pi)$ implies $H_{12}(\pi) = H_{23}(\pi)$. This means that $H_{ij}(\pi)$ splits $\Delta(X)$ in the same two regions regardless of i, j. Outside of such hyperplane, there will be no ties by construction. But this then means that one of the three alternatives never wins, contradicting FIE. Indeed, we can establish one of the many (similar) cases. For all $g \in \mathring{H}^+_{12}(\pi)$, a_1 beats a_2. Either $\mathring{H}^+_{12}(\pi) = \mathring{H}^+_{13}(\pi)$, or $\mathring{H}^+_{12}(\pi) = \mathring{H}^-_{13}(\pi)$. Consider the former. Then, in state θ with $f(\cdot|\theta) = g$, a_1 beats a_3. Also, either $\mathring{H}^+_{12}(\pi) = \mathring{H}^+_{23}(\pi)$ or $\mathring{H}^+_{12}(\pi) = \mathring{H}^-_{23}(\pi)$. Again, consider the former. Then a_2 beats a_3 at θ. For all $\hat{g}$ belonging to the other half-space and θ with $f(\cdot|\theta) = \hat{g}$, it has to be true that a_2 beats a_1, a_3 beats a_1, and a_3 beats a_2. But then a_2 does not win for almost every state, as we wanted to establish.

Move now to the *if* part. Consider a restricted 3-partition $\pi = \{E_1, E_2, E_3\}$ such that $P(\{\theta \in \Theta : f(\,\cdot\,|\theta) \in \mathcal{A}_i^\Delta \setminus \mathring{E}_i\}) = 0$ for $i = 1, 2, 3$. The are two distinct cases to consider.

Case 1: Not all E_i's share a facet. In this case, let i, j be such that $F_{ij}(\pi) = \emptyset$. Say it is $i = 1$ and $j = 3$, so E_1 and E_2 do share a facet. Let h_{12} be the normal of the hyperplane separating these two sets. Without loss, let $\mathring{H}^+(h_{12}) = \mathring{E}_1$. Similarly, let h_{23} denote the normal of the hyperplane separating E_2 and E_3 such that $\mathring{H}^-(h_{23}) = \mathring{E}_3$. We now construct a strategy σ that achieves FIE. Choose measurable functions $\hat{\sigma}_i : X \to \mathbb{R}_+$ such that $h_{12}(x) = \hat{\sigma}_1(x) - \hat{\sigma}_2(x)$ and $h_{23}(x) = \hat{\sigma}_2(x) - \hat{\sigma}_3(x)$. Choose $\varepsilon > 0$ sufficiently small such that $\sum_i (\hat{\sigma}_i(x)\varepsilon \leq 1$ for every x. Let $3R(x) = 1 - \sum_i \hat{\sigma}_i(x)\varepsilon$, and set $\sigma_i(x) = \hat{\sigma}_i(x)\varepsilon + R(x)$, so that $\sum_i \sigma_i(x) = 1$ for every $x \in X$. FIE now follows from simple computations. For instance, for almost all $\theta \in \mathcal{A}_1$, $z_1^\sigma(\theta) > z_2^\sigma(\theta)$ and also $z_2^\sigma(\theta) > z_3^\sigma(\theta)$, because $f(\cdot|\theta)$ lies on $\mathring{H}^+(h_{23})$.

Hence, a_1 wins for almost all $\theta \in \mathcal{A}1$. Similar computations establish that a_2 wins for almost all $\theta \in \mathcal{A}_2$ and a_3 wins for almost all $\theta \in \mathcal{A}_3$, so FIE is verified.

Case 2: All E_i's share a facet, so there is no pair i, j such that $F_{ij}(\pi) = \emptyset$. Let h_{ij} be the normal of a hyperplane separating E_i and E_j such that $\mathring{E}_i \subset \mathring{H}^+(h_{ij})$. Because π is a restricted 3-partition, it must be true that for all i, j, m, $H(h_{ij}) \neq H(h_{im})$. It must also be true that for all i, j, m, $H(h_{ij}) \cap H(h_{jm}) \subset H(h_{im})$. Indeed, suppose the inclusion does not hold, so we have $g \in H(h_{12}) \cap H(h_{23})$ and $g \notin H(h_{13})$. We can then find $\varepsilon > 0$ such that $B_\varepsilon(g) \cap H(h_{13}) = \emptyset$. Then, either $B_\varepsilon(g) \subset E_1 \cup E_2$ or $B_\varepsilon(g) \subset E_2 \cup E_3$. Suppose it is the latter. Because $g \in H(h_{12}) \cap H(h_{23})$, the ball $B_\varepsilon(g)$ has four regions formed by the intersection of half-spaces. In particular, either $\hat{E}_2 = \{\hat{g} \in B_\varepsilon(g) : \int h_{12}(x)\hat{g}(x)\lambda(dx) < 0 < \int h_{23}(x)\hat{g}(x)\lambda(dx)\}$ is strictly convex and $\hat{E}_3 = B_\varepsilon(g) \setminus E$ is not convex, or the other way around. Since $\hat{E}_i$ not convex implies E_i is not convex, we have a contradiction. As the choice of labels is arbitrary, we conclude that, for all i, j, m, $H(h_{ij}) \cap H(h_{jm}) \subset H(h_{im})$, and hence that the intersection of any two hyperplanes is the same. A hyperplane is a subspace of co-dimension 1 and the intersection of two hyperplanes is a subspace of co-dimension 2. Hence, the two-dimensional subspaces generated by the normals (h_{12}, h_{13}), (h_{12}, h_{23}), and (h_{13}, h_{23}) are the same. Hence, re-labeling if necessary, we can find scalars a and b such that $h_{13} = ah_{12} + bh_{23}$. Switching signs if necessary, it is without loss to have a and b strictly positive. Now set $\alpha = a/(a+b)$ and $\tilde{h}_{13} = h_{13}/(a+b)$ to establish that $\tilde{h}_{13}(x) = \alpha h_{12}(x) + (1-\alpha)h_{23}(x)$ for every $x \in X$. Observe that $\tilde{h}_{13}$ generates the same hyperplane as h_{13}. Now pick bounded measurable $\hat{\sigma}_i : X \to \mathbb{R}_+$ such that $\alpha h_{12}(x) = \hat{\sigma}_1(x) - \hat{\sigma}_2(x)$ and $(1-\alpha)h_{23}(x) = \hat{\sigma}_2(x) - \hat{\sigma}_3(x)$ for every x. As in Case 1 above, normalize $\hat{\sigma}$ to construct a strategy σ, and similar simple computations establish that σ achieves FIE.

8.3 Proof of Theorem 2

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

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

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, using property **PS**, consider $\theta \in \mathcal{A}_j$ such that $c_j^{-1} < \int h(x) f(x|\theta)\lambda(dx) < c_{j-1}^{-1}$. We want to show that $\int(\sigma_j(x) - \sigma_l(x))P(dx|\theta) > 0$ for all $l \neq j$.[16] 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 $l > j$. As $\int h(x)P(dx|\theta) > c_j^{-1}$, 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 $l < j$. As $\int h(x)P(dx|\theta) < c_{j-1}^{-1}$. 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

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

As we can apply the argument above for P-a.e. $\theta \in A_j$, FIE is verified.

8.4 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 $\sum_{\ell=1}^{L} f(x|\theta_\ell)\hat{\alpha}_\ell = f(x|\theta_1)$,

[16] Observe that here and in the rest of the argument we will not make use of the densities $f(\cdot|\cdot)$, so strictly speaking the result is true even when there's no underlying probability measure λ.

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 independence, for each m there is $h^m \in L_\infty(\lambda)$ (in fact, we can choose h^m to have range in $[-1, 1]$) such that

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

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. As $\mathcal{A}_i^m \subset \mathcal{A}_i^{m'}$ for $m' > m$, for each $\theta \in \mathcal{A}_i^m$ we have

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

As $\mathcal{A}_i^m$ converges to $\mathcal{A}_i$, for each such $\theta \in A_i$ we must also have $\int h(x) f(x|\theta)\lambda(dx) = \hat{c}_i$. Indeed, there must exist a sequence θ^m with $\theta^m \in \mathcal{A}_i^m$ such that $\theta^m \to \theta$. As $f(x|\theta^m) \to f(x|\theta)$ for each $x \in X$, by Lebesgue Dominated Convergence we have

$$\int h(x) f(x|\theta^m)\lambda(dx) \to \int h(x) f(x|\theta)\lambda(dx).$$

Property **PS** is therefore verified.

8.5 Proof of Lemma 3

Let $I = H(h_{ij}) \cap H(h_{jm})$. Suppose first that I is non-empty. If $f(\cdot|\theta) \in I$ then $\tilde{u}(\theta, a_i) = \tilde{u}(\theta, a_j) = \tilde{u}(\theta, a_m)$, hence $f(\cdot|\theta) \in H(h_{im})$. As in the proof of Theorem 1, there are scalars α_{ij} and α_{jm} such that $h_{im} = \alpha_{ij} h_{ij} + \alpha_{jm} h_{jm}$. Suppose instead I is empty. This is true whenever $H(h_{ij})$ and $H(h_{jm})$ are parallel. It is clear that there are 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 θ such that

$$\int h_{ij}(x) f(x|\theta)\lambda(dx) > 0 \text{ and } \int h_{jm}(x) f(x|\theta)\lambda(dx) > 0,$$

or there exists θ such that

$$\int h_{ij}(x)f(x|\theta)\lambda(dx) < 0 \text{ and } \int h_{im}(x)f(x|\theta)\lambda(dx) < 0.$$

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, $\hat{u}(\theta, a_i) > \hat{u}(\theta, a_j)$ and $\hat{u}(\theta, a_j) > \hat{u}(\theta, a_m)$. By transitivity, $\hat{u}(\theta, a_i) > \hat{u}(\theta, a_m)$. On the other hand,

$$\int (\alpha_{ij}h_{ij}(x) + \alpha_{jm}h_{jm}(x))f(x|\theta)\lambda(dx) < 0 \text{ and thus } \int h_{im}(x)f(x|\theta)\lambda(dx) < 0,$$

so that $\hat{u}(\theta, a_i) < \hat{u}(\theta, a_m)$, which is a contradiction.

Case 2. Suppose $\alpha_{ij} > 0$, and $\alpha_{jm} < 0$. Consider θ such that

$$\int h_{im}(x)f(x|\theta)\lambda(dx) = 0 \text{ and } \int h_{ij}(x)f(x|\theta)\lambda(dx) \neq 0.$$

The former implies that

$$\int (\alpha_{ij}h_{ij}(x) + \alpha_{jm}h_{jm}(x))f(x|\theta)\lambda(dx) = 0,$$

and hence

$$\int \alpha_{ij}h_{ij}(x)f(x|\theta)\lambda(dx) = -\int \alpha_{jm}h_{jm}(x)f(x|\theta)\lambda(dx).$$

There are two possibilities: (i) $\int h_{ij}(x)f(x|\theta)\lambda(dx) > 0$, which implies $\int h_{jm}(x)f(x|\theta)\lambda(dx) > 0$ and hence $\hat{u}(\theta, a_i) > \hat{u}(\theta, a_j) > \hat{u}(\theta, a_m)$; (ii) $\int h_{ij}(x)f(x|\theta)\lambda(dx) < 0$, which implies $\int h_{jm}(x)f(x|\theta)\lambda(dx) < 0$ and hence $\hat{u}(\theta, a_i) < \hat{u}(\theta, a_j) < \hat{u}(\theta, a_m)$. In either case we have a contradiction because $\int h_{im}(x)f(x|\theta)\lambda(dx) = 0$ implies that $\hat{u}(\theta, a_i) = \hat{u}(\theta, a_m)$.

Case 3. Suppose $\alpha_{ij} < 0$, and $\alpha_{jm} > 0$. This is symmetric to the case 2, so it cannot happen.

Finally, set $\alpha_{im} = 1$ and the proof is complete.

8.6 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 $h_{jl} = ch_{il}$, 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 θ such that $\hat{u}(\theta, a_i) = \hat{u}(\theta, a_j) = \hat{u}(\theta, a_l)$ but the second inequality implies that $\hat{u}(\theta, a_i) \neq \hat{u}(\theta, a_j)$ for all θ.

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.

As in the proofs of Theorems 1 and 2, we can now normalize τ to yield a 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 $j \neq 1$,

$$\begin{aligned} u(\theta, a_1) > u(\theta, a_j) \quad &\Longleftrightarrow \quad \int h_{1j}(x) f(x|\theta)\lambda(dx) > 0 \\ &\Longleftrightarrow \quad \int (\tau_1(x) - \tau_j(x)) f(x|\theta)\lambda(dx) > 0 \\ &\Longleftrightarrow \quad \int (\sigma_1(x) - \sigma_j(x)) f(x|\theta)\lambda(dx) > 0. \end{aligned}$$

Thus, a_1 obtains more votes than any alternative a_j. The same relationship holds with weak inequality and equality.

Consider now the case where alternative $a_i \neq a_1$ is the best alternative. Then

$$\begin{aligned} u(\theta, a_i) > u(\theta, a_j) \quad &\iff \quad \int h_{ij}(x) f(x|\theta)\lambda(dx) > 0 \\ &\iff \quad \int (\alpha_{1j} h_{1j}(x) - \alpha_{1i} h_{1i}) f(x|\theta)\lambda(dx) > 0 \\ &\iff \quad \int (\tau_1(x) - \tau_j(x) - \tau_1(x) + \tau_i(x)) f(x|\theta)\lambda(dx) \\ &= \quad \int (\tau_i(x) - \tau_j(x)) f(x|\theta)\lambda(dx) > 0 \\ &\iff \quad \int (\sigma_i(x) - \sigma_j(x)) f(x|\theta)\lambda(dx) > 0. \end{aligned}$$

Thus, a_i obtains more votes than any alternative a_j.Therefore, the strategy always chooses the right alternative in the limit and this concludes the proof.

8.7 Proof of Proposition 1

Without loss, we can take $u_a \in L_1^{\Delta}(\lambda)$ for each $a \in A$, by taking positive affine transformations if needed. Let σ be given by $\sigma_i(x) = u_{a_i}(x)$, so that $u(f, a_i) > u(f, a_j)$ if and only if $z_i^{\sigma}(f) > z_j^{\sigma}(f)$, so σ achieves FIE. For the converse, simply set $\hat{u}_{a_i} = \sigma_i$ for every a_i.

8.8 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. 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, $P(\mathcal{A}_j \backslash \mathcal{A}_j^{\sigma^\infty}) > 0$ 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.

8.9 Proof of Proposition 2

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.

8.10 Proof of Theorem 5

Fix $V \in \mathbb{N}_+$. 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.

First, notice that the strategy set under plurality rule is a subset of that under any V-scoring rule without restriction, hence whenever FIE is achieved under plurality rule, it is also achieved under a V-scoring rule without restriction.

Next, we show that whenever FIE is achieved under a V-scoring rule with or without restrictions, it is also achieved under the plurality rule. 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

$$\int \sigma_j^{\text{sum}}(x) P(dx|\theta).$$

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

$$\int \sigma_i^{\text{sum}}(x) P(dx|\theta) - \int \sigma_j^{\text{sum}}(x) P(dx|\theta)$$

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

$$\int \sigma_i^{\text{PV}}(x) P(dx|\theta) - \int \sigma_j^{\text{PV}}(x) P(dx|\theta).$$

This latter 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.