Epistemic democracy with correlated voters

Pivato, Marcus

THEMA, Université de Cergy-Pontoise, France

9 June 2017

Epistemic democracy with correlated voters

Marcus Pivato
THEMA, Université de Cergy-Pontoise*

June 9, 2017

Abstract

We develop a general theory of epistemic democracy in large societies, which subsumes the classical Condorcet Jury Theorem, the Wisdom of Crowds, and other similar results. We show that a suitably chosen voting rule will converge to the correct answer in the large-population limit, even if there is significant correlation amongst voters, as long as the *average* covariance between voters becomes small as the population becomes large. Finally, we show that these hypotheses are consistent with models where voters are correlated via a social network, or through the DeGroot model of deliberation.

Keywords: Condorcet Jury Theorem; Wisdom of Crowds; epistemic social choice; deliberation; social network; DeGroot.
JEL class: D71, D81.

1 Introduction

The epistemic approach to social choice theory originates with Condorcet (1785). Suppose a group of people want to obtain the correct answer to some dichotomous (yes/no) question. The question has an objectively correct answer, and everyone has an opinion, but nobody has perfect information. The group could be, for example, a jury trying to determine the guilt or innocence of a defendant in a criminal trial, or a committee of engineers trying to determine whether a bridge is structurally safe. Condorcet's insight was that such a group could efficiently aggregate their private information by *voting*. He assumed that each voter's success rate at divining the truth was better than blind chance, and that the errors of different voters were stochastically independent. The famous Condorcet Jury Theorem (CJT) then consists of two statements:

- A decision made by a committee using majority vote will be more reliable than the opinion of any single individual. Furthermore, larger committees are more reliable than smaller committees.

*33 Boulevard du Port, 95011 Cergy-Pontoise cedex, France. email: `marcuspivato@gmail.com`

- Majority vote will converge in probability to the correct answer as the committee size becomes arbitrarily large.

The first statement is sometimes called the *nonasymptotic* part of the CJT, while the second statement is the *asymptotic* part. Although it was originally stated only for dichotomous decisions made by majority vote, the CJT has been generalized to polychotomous decisions made by the plurality rule (Lam and Suen, 1996; Ben-Yashar and Paroush, 2001; List and Goodin, 2001), and even other voting rules such as the Kemeny rule and the Borda rule (Young, 1986, 1988, 1995, 1997). Furthermore, in these contexts, the "nonasymptotic" part of the CJT can be refined: under certain conditions, the output of the voting rule is a *maximum likelihood estimator* of the correct answer (see Pivato (2013b) for a general formulation of these results).

A closely related result is the "Wisdom of Crowds" (WoC) principle of Galton (1907): if a large number of people independently estimate some numerical quantity, then the average of their estimates will converge, in probability, to the true value. However, the WoC, the CJT, and all of its polychotomous generalizations depend on the assumption that the errors made by different voters are *independent* random variables. This is obviously unrealistic: in reality, the opinions of different voters will be strongly correlated, both because they rely on common sources of information and because they influence one another through deliberation and discussion. The goal of this paper is to extend the asymptotic part of the CJT, WoC, and similar theorems to an environment with *correlated* voters.

It has been understood for a long time that the "independence" assumption in the CJT is problematic. Starting in the 1980s, a series of papers gauged the seriousness of this problem and proposed possible solutions. Nitzan and Paroush (1984) demonstrated the sensitivity of the CJT to the independence assumption, while Shapley and Grofman (1984) showed that, with certain patterns of correlations, a *nonmonotonic* rule could actually be more reliable than majority vote. Owen (1986) argued that, if the voters can be divided into subgroups such that voters within each subgroup are correlated, then an "indirect" majority vote (like an electoral college) could be more reliable than direct majority vote. Meanwhile, Ladha (1992) showed that the asymptotic CJT remained true as long as the "average" covariance between the voters was sufficiently small. (This is a special case of Theorem 5.3 in the present paper.) Berend and Sapir (2007) found general conditions for the nonasymptotic part of the CJT to hold in a committee of correlated voters. Kaniovski (2009, 2010) modeled the joint probability distribution of a population of homogeneous correlated voters using a representation by Bahadur, and studied the nonasymptotic part of the CJT in this context. Building on this work, Kaniovski and Zaigraev (2011) showed that a special case of the Bahadur representation admits a quota voting rule which neutralizes the effect of the correlations. Finally, Peleg and Zamir (2012) gave a number of necessary conditions and sufficient conditions for a population of correlated voters to satisfy the CJT.

One natural source of voter correlation is "contagion" of opinions (e.g. due to deliberation). Berg (1993a,b) and Ladha (1995) supposed that the voters' errors were correlated according to hypergeometric or Pólya-Eggenberger urn processes, which are simple models of such "contagion". They showed that the asymptotic CJT holds for the former, but does not hold for the latter (although a group is still more reliable than an individual). See Berg (1996) for a summary.

Another possible cause of voter correlation is a common source of information. For example, in a criminal trial, all jurors observe exactly the same evidence (although they may interpret it differently). In a committee of engineers, everyone reads the same technical reports and has access to the same data. In other situations, the voters might all be influenced by an "opinion leader". Boland (1989) and Boland et al. (1989) developed a version of the CJT with such an opinion leader. Later, Berg (1994) extended this to a setting with weighted voting rules. Estlund (1994) also considered a model with opinion leaders, but he showed that, under certain conditions, such opinion leaders could actually improve the reliability of majority vote. Meanwhile, Ladha (1993, Proposition 1) proved a version of the CJT when the voter errors are not independent, but are *exchangeable* random variables. By a theorem of de Finetti, this is equivalent to a model where the voter opinions are independent Bernoulli random variables with a common parameter α, which is itself another random variable; thus, α can be interpreted as representing a common information source. (The Pólya-Eggenberger distributions studied by Berg (1993a,b) and Ladha (1995) are also examples of exchangeable distributions.) Peleg and Zamir (2012, Theorem 5) also proved a version of the CJT for exchangeable random variables. Dietrich and List (2004) demonstrated that if all voters draw only on a small set of (unreliable) information sources, then the asymptotic part of the CJT fails: even a very large population of voters cannot be any more reliable than the (small) set of information sources on which they all base their opinions. Dietrich and List represented this situation as a Bayesian network; this approach was further developed by Dietrich and Spiekermann (2013a,b), who showed that, in the presence of common causes, the asymptotic reliability of a large committee can be good, but less than perfect.

A third possible cause of correlation is strategic voting. Even if all voters want the group to get the correct answer, they may have incentives to vote strategically (Austen-Smith and Banks, 1996). To see this, recall that each voter's optimal voting strategy is based on the hypothesis that she is a pivotal voter. But this hypothesis has implications for how other people must have voted, and thus, indirectly, about the state of the world itself. So a voter who believes that she is pivotal has extra information beyond her private signal, and this may change the way she votes; in some cases, she may actually vote *against* her private information. But in a strategic setting, *all* voters will vote "as if" they are pivotal, so such strategic dishonesty may be widespread (and correlated) in equilibrium.

However, the consequences of strategic voting are not as dire as one might imagine in an epistemic context. McLennan (1998, Theorem 1) has shown that any profile of voting strategies which maximizes the probability that the group gets the right answer will be a Bayesian Nash equilibrium (BNE). This holds even if the voters' types (i.e. their private information) are correlated. As observed by Peleg and Zamir (2012), this means that we only need to prove the existence of *some* pattern of voting behaviour which satisfies the CJT; it then follows that the CJT will also hold in BNE. Thus, we do not need to explicitly consider strategic behaviour in our analysis.

Aside from voter correlation, another important issue in epistemic social choice theory is the tradeoff between group size and average voter competency. Suppose we could arrange the voters in order from most epistemically competent to least competent. We could then consider various forms of "epistocracy", where we delegate the decision to the N most competent individuals, for some value of N. One extreme ($N = 1$) is rule by a "philosopher king" —the single most competent individual. The opposite extreme (maximal N) is "mass democracy", where all voters participate equally. If all voters are equally competent, then Condorcet's theorem says that increasing N will alway leads to better decisions. But we can easily imagine situations where competency is distributed so unequally amongst the voters that increasing N will lead to *worse* decisions. The problem is exacerbated if the competency of each individual voter is itself a decreasing function of the size of the electorate in which she participates. This is plausible if, for example, there is a fixed budget of resources to spend on educating and informing the voters (so that increasing N necessarily decreases the educational resources available for each voter), or if voters in a larger electorate are tempted to epistemically "free ride" on their colleagues. A series of papers have considered this size/competency tradeoff (Boland, 1989; Kanazawa, 1998; Karotkin and Paroush, 2003; Berend and Sapir, 2007). Our results show that the asymptotic results of the CJT and WoC remain true even if average voter competency decreases as the population size increases —as long as it does not decay *too* quickly.

Almost all of the aforementioned papers deal only with *dichotomous* decision problems and majority rule.[1] In contrast, the asymptotic results of this paper are applicable to *polychotomous* decisions and a large class of epistemic voting rules, including majority rule, plurality rule, the Kemeny rule, the median rule (on a discrete metric space), the average rule (for vector-valued decisions), Condorcet-consistent rules, and scoring rules such as the Borda rule. To obtain this level of generality, we will introduce a single broad class of voting rules which includes all of the aforementioned rules as special cases: the class of *mean partition rules*. This class of rules yields a very general approach to epistemic social choice theory, which subsumes all existing versions of the asymptotic CJT (dichotomous and polychotomous) and the WoC principle, and also applies to many other epistemic social choice models. We will show that these asymptotic results can remain valid even when there is considerable correlation between voters, and even if the average competency of voters decreases as the population increases. Furthermore, we will provide a concrete illustration of the economic relevance of our general results, by connecting them with the theory of social networks and with the DeGroot (1974) model of consensus formation.

The remainder of this paper is organized as follows. Section 2 introduces notation and terminology which will be maintained throughout the paper. Section 3 defines the class of mean partition rules and gives several examples, including majority rule, plurality rule, the median rule, and other scoring rules. Section 4 describes a special case of our model, which we call a *populace*: this is a family of probability distributions, describing a society where voters make *independent* random errors. It contains two special cases of

[1] Exceptions are Young (1986, 1988, 1995, 1997), Lam and Suen (1996), McLennan (1998), Ben-Yashar and Paroush (2001), and List and Goodin (2001).

our main result, which state that, if the populace satisfies certain mild conditions, then an appropriate mean partition rule will select the correct answer with very high probability in a large population (Propositions 4.1 and 4.3).

Section 5 describes the general model, which we call a *culture*: this is a family of probability distributions, describing a society where the errors of the voters are *correlated* random variables. It then states the general version of our main result (Theorem 5.3): if the culture is *sagacious* (meaning that it satisfies certain mild conditions —in particular, the "average covariance" between voters is not too large), then an appropriate mean partition rule will select the correct answer with very high probability in a large population.

The rest of the paper explores applications. Section 6 considers cultures based on social networks, and contains results (Propositions 6.2 and 6.5) stating that, as long as the social network is not too richly connected, the resulting culture will be sagacious, so that Theorem 5.3 applies. Finally, Section 7 considers the effects of deliberation on an already sagacious culture, and contains a result (Proposition 7.1) saying that, as long as no individuals are too "influential" in this deliberation, the culture will remain sagacious after deliberation. All proofs are in the Appendix.

2 Notation and terminology

We now fix some notation which will be maintained throughout the paper. Let $\mathbb{N} := \{1, 2, \ldots\}$ denote the set of natural numbers. Let $\mathbb{R}$ denote the set of real numbers, and let $\mathbb{R}_+$ denote the set of nonnegative real numbers. Let $\mathcal{I}$ denote a finite set of voters, and let $I := |\mathcal{I}|$. (We will typically assume that I is very large; indeed, we will mainly be interested in asymptotic properties as $I \to \infty$.)

A ***metric space*** is a set $\mathcal{S}$ together with a function $d : \mathcal{S} \times \mathcal{S} \longrightarrow \mathbb{R}_+$ such that, for any $r, s, t \in \mathcal{S}$: (1) $d(s,t) = d(t,s)$; (2) $d(s,t) = 0$ if and only if $s = t$; and (3) $d(r,t) \leq d(r,s) + d(s,t)$. We will assume throughout the paper that the set of possible states of the world is a metric space (for example, a subset of some Euclidean space). If $\mathcal{S}$ is a finite set, then we will just use the discrete metric, where $d(s,t) = 1$ for any $s \neq t$.

If $\mathbb{V}$ is a vector space, then a ***norm*** on $\mathbb{V}$ is a function $\|\_\| : \mathbb{V} \longrightarrow \mathbb{R}_+$ such that, for any $\mathbf{v}, \mathbf{w} \in \mathbb{V}$ and $r \in \mathbb{R}$: (1) $\|r\,\mathbf{v}\| = |r| \cdot \|\mathbf{v}\|$; (2) $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = 0$; and (3) $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$. Such a norm defines a metric d on $\mathbf{V}$ by $d(\mathbf{v}, \mathbf{w}) := \|\mathbf{v} - \mathbf{w}\|$. An ***inner product*** on $\mathbb{V}$ is a function $\langle \_, \_ \rangle : \mathbb{V} \times \mathbb{V} \longrightarrow \mathbb{R}$ such that, for any $\mathbf{v}, \mathbf{w} \in \mathbb{V}$: (1) The functions $\langle \mathbf{v}, \_ \rangle : \mathbb{V} \longrightarrow \mathbb{R}$ and $\langle \_, \mathbf{w} \rangle : \mathbb{V} \longrightarrow \mathbb{R}$ are linear; (2) $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$; and (3) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and furthermore, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = 0$. For example, if $\mathbb{V} = \mathbb{R}$, then we could simply take $\langle r, s \rangle := r\,s$ for any $r, s \in \mathbb{R}$. If $\mathbb{V} = \mathbb{R}^N$, then we could use the standard dot product: $\langle \mathbf{v}, \mathbf{w} \rangle = v_1\,w_1 + \cdots + v_N\,w_N$ for any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^N$. An inner product defines a norm by setting $\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. For example, the ***Euclidean norm*** on $\mathbb{R}^N$ is defined by $\|\mathbf{v}\| := \sqrt{v_1^2 + \cdots + v_N^2}$. An ***inner product space*** is a vector space equipped with an inner product. We always endow such a space with the metric induced by the norm induced by the inner product. We will assume throughout the paper that the set of votes that can be sent by the voters is a subset of some inner product space.

Let ρ be a probability measure on a vector space $\mathbb{V}$. The ***expected value*** of a ρ-random

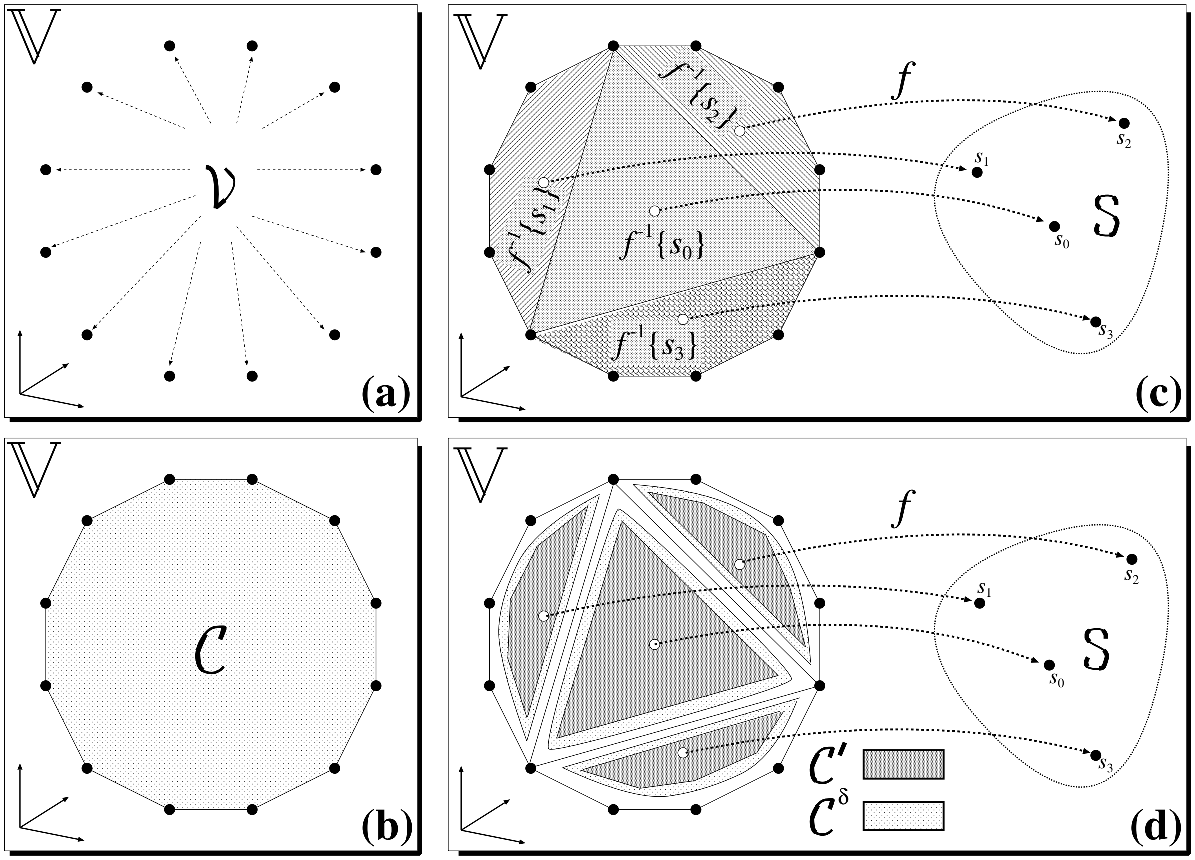

Figure 1: A mean partition rule. **(a)** $\mathcal{V}$ is a subset of the vector space $\mathbb{V}$. **(b)** $\mathcal{C}$ is the convex hull of $\mathcal{V}$. **(c)** $f^{-1}\{s\}$ is a convex subset of $\mathcal{C}$, for each $s \in \mathcal{S}$. **(d)** The sets $\mathcal{C}'$ and $\mathcal{C}^\delta$.

variable is defined $\mathbb{E}[\rho] := \int_{\mathcal{V}} \mathbf{v} \; \mathrm{d}\rho[\mathbf{v}]$. If $\mathbb{V}$ has a norm $\|.\|$, then the variance is defined $\mathrm{var}[\rho] := \displaystyle\int_{\mathcal{V}} \|\mathbf{v} - \overline{\mathbf{v}}\|^2 \; \mathrm{d}\rho[\mathbf{v}]$, where $\overline{\mathbf{v}} := \mathbb{E}[\rho]$.

3 Mean partition rules

This section introduces *mean partition rules*: voting rules where the outcome is functionally determined by the *average* of the signals sent by the voters. After formally defining this class of rules, we provide a series of examples, showing that many common voting rules fall into this class.

Let $\mathcal{I}$ be a set of individuals. Let $\mathcal{S}$ be a metric space, whose elements represent social alternatives. An (anonymous) *mean partition rule* on $\mathcal{S}$ is a voting rule defined by a data structure $F := (\mathbb{V}, \mathcal{V}, f)$ with four properties:

(M1) $\mathbb{V}$ is an inner product space, and $\mathcal{V} \subseteq \mathbb{V}$ (as shown in Figure 1(a)).

(M2) Let $\mathcal{C}$ be the closed convex hull of $\mathcal{V}$ (as in Figure 1(b)).[2] Then $f : \mathcal{C} \longrightarrow \mathcal{S}$ is a surjective function (as shown in Figure 1(c)).

(M3) There exists a subset $\mathcal{C}' \subseteq \mathcal{C}$ and $\delta > 0$ such that, if we define $\mathcal{C}^\delta := \{\mathbf{c} \in \mathcal{C};\ d(\mathbf{c}, \mathcal{C}') < \delta\}$, then f is uniformly continuous and surjective when restricted to $\mathcal{C}^\delta$.

(M4) For any $s \in \mathcal{S}$, the set $f^{-1}\{s\} \cap \mathcal{C}'$ is convex (as in Figure 1(d)).

In this model, $\mathcal{V}$ is the set of possible ***votes*** which could be sent by each individual. Given any finite set $\mathcal{I}$ of individuals, and any profile $\mathbf{V} = (\mathbf{v}_i)_{i\in\mathcal{I}}$ of votes (where $\mathbf{v}_i \in \mathcal{V}$ for all $i \in \mathcal{I}$), the output of the rule F is obtained by applying f to the *average* of the vectors $\{\mathbf{v}_i\}_{i\in\mathcal{I}}$. Formally,

$$F(\mathbf{V}) \quad := \quad f\left(\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \mathbf{v}_i\right), \qquad \text{for all } \mathbf{V} = (\mathbf{v}_i)_{i\in\mathcal{I}} \in \mathcal{V}^{\mathcal{I}}. \tag{1}$$

A few remarks are in order. First, if $\mathbb{V}$ is a *finite*-dimensional vector space, then it always has an inner product, and furthermore all inner products on $\mathbb{V}$ are uniformly equivalent; thus, the requirement that $\mathbb{V}$ be an inner product space in (M1) involves absolutely no loss of generality. Second, from property (M2) and equation (1), it is clear that the voting rule F is anonymous by construction (i.e. the outcome is invariant under permutation of the voters). Third, (M3) does *not* require f to be continuous everywhere on $\mathcal{C}$. (Indeed, if $\mathcal{S}$ were a discrete set, this would be impossible.) However, if f is injective (so that $f^{-1}\{s\}$ is a singleton for all $s \in \mathcal{S}$), then the surjectivity part of (M3) implies that $\mathcal{C}^\delta = \mathcal{C}$, so that f is a uniformly continuous function on $\mathcal{C}$. (In this case, the convexity condition (M4) is automatically satisfied.) At the other extreme, if $\mathcal{S}$ is finite, then (M2) says that f defines an $\mathcal{S}$-labelled partition of $\mathcal{C}$ —in other words, $\mathcal{C} = \bigcup_{s\in\mathcal{S}} \mathcal{C}_s$, where $\mathcal{C}_s := f^{-1}\{s\}$ for all $s \in \mathcal{S}$. Meanwhile, (M4) says that $\mathcal{C}' = \bigcup_{s\in\mathcal{S}} \mathcal{C}'_s$, where $\mathcal{C}'_s$ is a convex subset of $\mathcal{C}_s$, for each $s \in \mathcal{S}$. Figure 1(c) suggests that $\mathcal{C}_s$ is *also* convex for each $s \in \mathcal{S}$, and indeed, this is the case in many of our examples. But it is not true in general. Since f is single-valued, it must use some "tie-breaker" rules for points along the boundaries between the preimage sets $\mathcal{C}_s$ (for $s \in \mathcal{S}$), and the sets $\mathcal{C}_s$ would be convex only if these tie-breaker rules were carefully chosen. Fortunately, this doesn't matter; the sets $\mathcal{C}_s$ *need not be convex*, as long as (M4) is satisfied. (See Example 3.3(a) below for an illustration.) Indeed, it is for precisely this reason that (M3) introduced $\mathcal{C}'$.[3]

Example 3.1. (a) (*Simple majority rule*) Let $\mathcal{S} := \{\pm 1\}$. Let $\mathbb{V}_{\text{maj}} := \mathbb{R}$. Let $\mathcal{V}_{\text{maj}} := \{\pm 1\}$, so that $\mathcal{C} = [-1, 1]$, as shown in Figure 2(a). Define $f_{\text{maj}} : \mathcal{C} \longrightarrow \mathcal{S}$ by setting $f_{\text{maj}}(r) := \text{sign}(r)$ for all nonzero $r \in [-1, 1]$, while $f_{\text{maj}}(0) := 1$ (an arbitrary tie-breaking rule). Then $F_{\text{maj}} = (\mathbb{V}_{\text{maj}}, \mathcal{V}_{\text{maj}}, f_{\text{maj}})$ is the simple majority rule. Now, fix $\epsilon > 0$, and let $\mathcal{C}' := \mathcal{C}'_{-1} \sqcup \mathcal{C}'_{+1}$, where $\mathcal{C}'_{-1} := [-1, -\epsilon)$ and $\mathcal{C}'_{+1} := (\epsilon, 1]$, as shown in Figure 2(b). Then (M3) and (M4) are satisfied (set $\delta := \epsilon/2$).

[2] That is: $\mathcal{C}$ is the smallest closed, convex subset of $\mathbb{V}$ that contains $\mathcal{V}$. Equivalently, $\mathcal{C}$ is the intersection of all closed convex sets containing $\mathcal{V}$. If $\mathcal{V}$ is finite, then its convex hull is automatically closed, so in this case we could just define $\mathcal{C}$ to be its convex hull.

[3] I thank the referee for calling my attention to this issue.

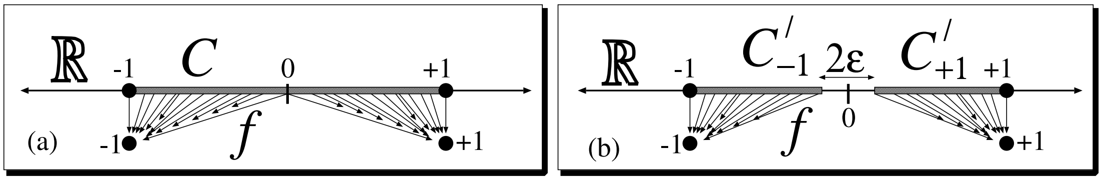

Figure 2: (Example 3.1(a)) Simple majority vote as a mean partition rule.

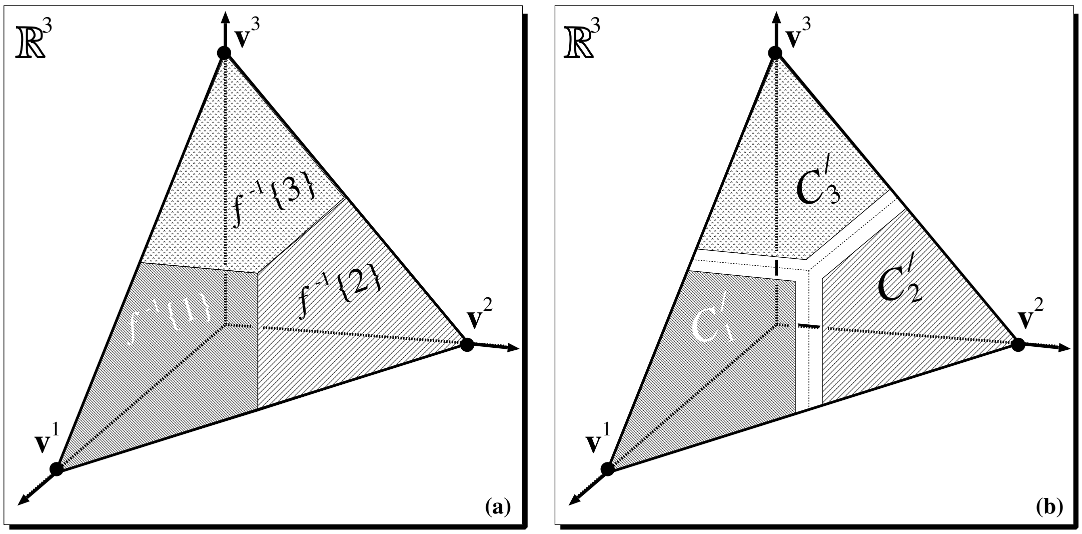

Figure 3: (Example 3.1(b)) The plurality rule as a mean partition rule.

Throughout the remaining examples, let $\wp(\mathcal{S})$ be the power set of $\mathcal{S}$, and let $\tau : \wp(\mathcal{S}) \longrightarrow \mathcal{S}$ be a function such that $\tau(\mathcal{Q}) \in \mathcal{Q}$ for all nonempty $\mathcal{Q} \subseteq \mathcal{S}$. (Thus, in particular, $\tau\{s\} = s$ for all $s \in \mathcal{S}$.) We will use τ as a "tie-breaker" in the definition of the following rules.

(b) (*Plurality rule*) Let $N \geq 2$, and let $\mathcal{S} := \{1, 2, \ldots N\}$ (a set of N alternatives). Let $\mathbb{V}_{\text{plu}} := \mathbb{R}^N$. For all $n \in [1 \ldots N]$, let $\mathbf{v}^n := (0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 appears in the nth coordinate. Let $\mathcal{V}_{\text{plu}} := \{\mathbf{v}^1, \ldots, \mathbf{v}^N\}$ (a subset of $\mathbb{R}^N$). If $\mathcal{C}$ is the convex hull of $\mathcal{V}$, then $\mathcal{C}$ is the unit simplex in $\mathbb{R}^N$, as shown in Figure 3(a). For any $\mathbf{c} \in \mathcal{C}$, let $\mathcal{S}_{\mathbf{c}} := \{s \in \mathcal{S};\ c_s \geq c_t$ for all $t \in \mathcal{S}\}$ be the set of maximal coordinates. Define $f_{\text{plu}} : \mathcal{C} \longrightarrow \mathcal{S}$ by setting $f_{\text{plu}}(\mathbf{c}) := \tau(\mathcal{S}_{\mathbf{c}})$, for all $\mathbf{c} \in \mathcal{C}$. Then $F_{\text{plu}} = (\mathbb{V}_{\text{plu}}, \mathcal{V}_{\text{plu}}, f_{\text{plu}})$ is the plurality rule. Fix $\epsilon \in (0, 1)$, and for all $s \in \mathcal{S}$, define the convex set $\mathcal{C}'_s := \{\mathbf{c} \in \mathcal{C};\ c_s > c_t + \epsilon$ for all $t \neq s\}$, as shown in Figure 3(b). (Note that $\mathcal{C}'_s \neq \emptyset$ because $\epsilon < 1$.) Let $\mathcal{C}' := \mathcal{C}'_1 \sqcup \mathcal{C}'_2 \sqcup \cdots \sqcup \mathcal{C}'_N$; then (M3) and (M4) are satisfied (set $\delta := \epsilon/2$).

(c) (*The median rule*) Let $\mathcal{S}$ be a finite subset of $\mathbb{R}$. Let $\mathbb{V}_{\text{med}} := \mathbb{R}^{\mathcal{S}}$. For all $s \in \mathcal{S}$, define $\mathbf{v}^s := (v^s_t)_{t \in \mathcal{S}} \in \mathbb{V}$, by setting $v^s_t := |s - t|$ for all $t \in \mathcal{S}$. Let $\mathcal{V}_{\text{med}} := \{\mathbf{v}^s\}_{s \in \mathcal{S}}$ (a subset of $\mathbb{V}_{\text{med}}$), and let $\mathcal{C}$ be the convex hull of $\mathcal{V}$. For any $\mathbf{c} \in \mathcal{C}$, let $\mathcal{S}_{\mathbf{c}} := \{s \in \mathcal{S};\ \ c_s \leq c_t$ for all $t \in \mathcal{S}\}$ be the set of minimal coordinates of $\mathbf{c}$ —in effect, these are the element(s) of $\mathcal{S}$ which minimize the *average distance* to the points chosen by the voters. It is easy to see that $\mathcal{S}_{\mathbf{c}}$ is always an interval inside $\mathcal{S}$. Define $f_{\text{med}} : \mathcal{C} \longrightarrow \mathcal{S}$ by setting $f_{\text{med}}(\mathbf{c}) := \tau(\mathcal{S}_{\mathbf{c}})$, for all $\mathbf{c} \in \mathcal{C}$. In other words, each voter chooses a point s in $\mathcal{S}$ (represented by $\mathbf{v}^s$), and

$F_{\rm med}$ chooses a point in $\mathcal{S}$ which minimizes the *average distance* to the points chosen by the voters (using τ to break ties). As is well-known, this point will be a *median* of the points chosen by the voters. (This is a special case of the next example.)

(d) (*The generalized median rule*) Let $(\mathcal{S}, d)$ be a finite metric space (for example, a connected graph with the geodesic metric). Let $\mathbb{V}_{\rm med} := \mathbb{R}^{\mathcal{S}}$. For all $s \in \mathcal{S}$, define $\mathbf{v}^s := (v^s_t)_{t\in\mathcal{S}} \in \mathbb{V}$, by setting $v^s_t := d(s,t)$ for all $t \in \mathcal{S}$. Let $\mathcal{V}_{\rm med} := \{\mathbf{v}^s\}_{s\in\mathcal{S}}$ (a subset of $\mathbb{V}_{\rm med}$), and let $\mathcal{C}$ be the convex hull of $\mathcal{V}_{\rm med}$. For any $\mathbf{c} \in \mathcal{C}$, let $\mathcal{S}_{\mathbf{c}} := \{s \in \mathcal{S};\ \ c_s \leq c_t$ for all $t \in \mathcal{S}\}$, as in example (d). Define $f_{\rm med} : \mathcal{C} \longrightarrow \mathcal{S}$ by setting $f_{\rm med}(\mathbf{c}) := \tau(\mathcal{S}_{\mathbf{c}})$, for all $\mathbf{c} \in \mathcal{C}$. As in example (d), each voter chooses a point s in $\mathcal{S}$ (represented by $\mathbf{v}^s$), and $F_{\rm med}$ selects a point in $\mathcal{S}$ which minimizes the *average distance* to the points chosen by the voters (using τ to break ties). To see that this is a mean partition rule, let $\epsilon > 0$, and for all $s \in \mathcal{S}$, let $\mathcal{C}'_s := \{\mathbf{c} \in \mathcal{C};\ \ c_s < c_t - \epsilon$ for all $t \in \mathcal{S} \setminus \{s\}\}$. If ϵ is small enough, then these sets are nonempty for all $s \in \mathcal{S}$, convex, and disjoint, and if we define $\mathcal{C}' := \bigsqcup_{s\in\mathcal{S}} \mathcal{C}'_s$ and $\delta := \epsilon/2$, then $f_{\rm med}$ is uniformly continuous on $\mathcal{C}'$; thus, (M3) and (M4) are satisfied.

(e) (*The Kemeny rule*) Let $\mathcal{A}$ be a finite set of social alternatives. Let $\mathcal{S}$ be the set of all linear orders over $\mathcal{A}$. The *Kendall metric* on $\mathcal{S}$ is defined by declaring $d(s,r)$ to be the number of pairwise comparisons where the orders s and r disagree. In this case, the generalized median rule from example (d) is the *Kemeny rule* for preference aggregation.

(f) (*Scoring rules*) Let $\mathcal{S}$ be a finite set of alternatives, and let $\mathbb{V}_{\rm scr} := \mathbb{R}^{\mathcal{S}}$. Let $\mathcal{V}$ be any subset of $\mathbb{V}$. Intuitively, an element $\mathbf{v} = (v_s)_{s\in\mathcal{S}}$ in $\mathcal{V}$ represents a vote which assigns a "score" of v_s to each alternative in $\mathcal{S}$. Let $\mathcal{C}$ be the convex hull of $\mathcal{V}$. For any $\mathbf{c} \in \mathcal{C}$, let $\mathcal{S}_{\mathbf{c}} := \{s \in \mathcal{S};\ \ c_s \geq c_t$ for all $t \in \mathcal{S}\}$ be the set of maximizers of $\mathbf{c}$. Define $f_{\rm scr} : \mathcal{C} \longrightarrow \mathcal{S}$ by setting $f_{\rm scr}(\mathbf{c}) := \tau(\mathcal{S}_{\mathbf{c}})$, for all $\mathbf{c} \in \mathcal{C}$. Then $F_{\rm scr} = (\mathbb{V}_{\rm scr}, \mathcal{V}, f_{\rm scr})$ is called a *scoring rule*. All of the examples above are special cases of scoring rules. Other well-known scoring rules include the *Borda rule* and the *Approval Voting* rule. (The proof that this is a mean partition rule is similar to example (d).)

(g) (*Mean proximity rules*) Let $\mathcal{S}$ be a finite set of alternatives, and for each $s \in \mathcal{S}$, let $\mathbf{r}_s \in \mathbb{R}^N$. Let $\mathcal{V}$ be another finite subset of $\mathbb{R}^N$. Let $\mathcal{C}$ be the convex hull of $\mathcal{V}$. For any $\mathbf{c} \in \mathcal{C}$, let $\mathcal{S}_{\mathbf{c}} := \{s \in \mathcal{S};\ \ \|\mathbf{r}_s - \mathbf{c}\|$ is minimal$\}$. Define $f_{\rm mpr} : \mathcal{C} \longrightarrow \mathcal{S}$ by setting $f_{\rm mpr}(\mathbf{c}) := \tau(\mathcal{S}_{\mathbf{c}})$, for all $\mathbf{c} \in \mathcal{C}$. Then $F_{\rm mpr} = (\mathbb{V}_{\rm scr}, \mathcal{V}, f_{\rm mpr})$ is called a *mean proximity rule*. $\diamondsuit$

The median rule in Example 3.1(c) might seem more like a statistical construct than a *bona fide* voting rule. But if all voters have single-peaked preferences on a linearly ordered set $\mathcal{S}$, then the median alternative is the Condorcet winner, so it will be the outcome of any Condorcet-consistent voting rule (Black, 1958). Medians also arise in another important voting rule. Let $\mathcal{A}$ be a set of alternatives, let $\mathcal{S}$ be a linearly ordered set of "rankings", and suppose each voter assigns an ranking in $\mathcal{S}$ to each alternative in $\mathcal{A}$. The *majority judgement* rule selects the alternative in $\mathcal{A}$ which receives the highest *median ranking* from the voters. This rule has many nice properties (Balinski and Laraki, 2011). Meanwhile, the generalized median voting rule of Example 3.1(d) has been studied and axiomatically characterized for finite metric spaces (Barthélémy and Janowitz, 1991), graphs and lattices (McMorris et al., 2000), and judgement aggregation (Nehring and Pivato, 2017).

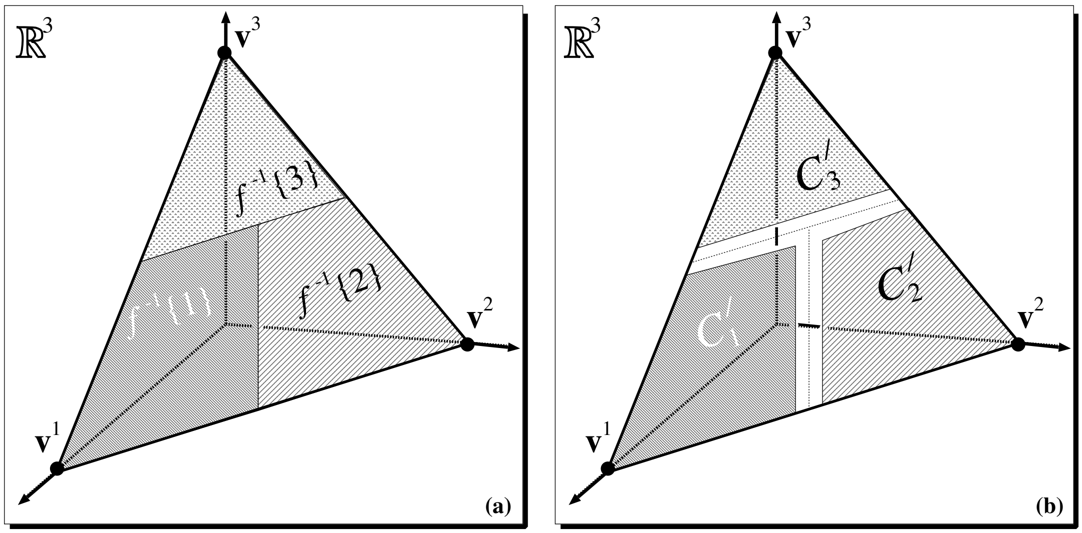

Figure 4: (Example 3.2) A (convex) mean partition rule that is not a scoring rule.

The scoring rules of Example 3.1(f) are related to the *generalized scoring rules* of Xia and Conitzer (2008). The difference is that Xia and Conitzer identify the elements of $\mathcal{V}$ with preference orders over $\mathcal{S}$; on the other hand, they do not necessarily use the maximizer as the winner. Xia (2015) introduced a further generalization he called *generalized decision scoring rules*, and proved a CJT-type result similar to Proposition 4.1 below. When $\mathcal{V}$ and $\mathcal{S}$ are both finite, Zwicker (2008, Theorem 4.2.1) has shown that an anonymous voting rule is a scoring rule (as in Example 3.1(f)) if and only if it is a mean proximity rule (as in Example 3.1(g)).[4] So these two classes are equivalent. But not every mean partition rule is a scoring rule, even when $\mathcal{V}$ and $\mathcal{S}$ are finite, as shown by the next example.

Example 3.2. (*Not a scoring rule*) Let $\mathcal{S} = \{1, 2, 3\}$, let $\mathbb{V} = \mathbb{R}^3$, and let $\mathcal{V} = \{\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3\}$, as in the *Plurality rule* of Example 3.1(b). Thus, $\mathcal{C}$ is the unit simplex in $\mathbb{R}^3$. Define $f : \mathcal{C} \longrightarrow \mathcal{S}$ as follows (see Figure 4(a)):

$$\text{for all } \mathbf{c} = (c_1, c_2, c_3) \in \mathcal{C}, \quad f(\mathbf{c}) \quad := \quad \begin{cases} 3 & \text{if} \quad c_3 > \frac{1}{2}; \\ 1 & \text{if} \quad c_3 \leq \frac{1}{2} \text{ and } c_1 \geq c_2; \\ 2 & \text{if} \quad c_3 \leq \frac{1}{2} \text{ and } c_1 < c_2. \end{cases}$$

Thus, alternative 3 wins if it is supported by a strict majority of the voters; otherwise either 1 or 2 wins, depending on which of them is supported by more voters (with ties broken in favour of alternative 1). For example, $f(0.3, 0.25, 0.45) = 1$. Figure 4(b) illustrates how this is a mean partition rule. But it is not a scoring rule (Pivato, 2013a, Example 2).[5]

Such a rule would make sense in a scenario where alternative 3 was seen as *prima facie* less desirable than alternatives 1 or 3, so that it needs a higher level of popular support

[4] Zwicker's model is slightly different: instead of using a tiebreaker rule, he allows voting rules to be multivalued in the case of a tie.

[5] Cervone and Zwicker (2009) contains a similar example, but their focus is on convex partitions rather than scoring rules.

to be adopted. In an epistemic context, alternative 3 might be regarded as less plausible than alternatives 1 or 2, and thus demanding a higher standard of evidence. ♢

In all the mean partition rules in Examples 3.1 and 3.2, the function f defines a convex, $\mathcal{S}$-labelled partition of the convex hull $\mathcal{C}$; the continuity set $\mathcal{C}'$ in condition (M3) is then obtained by "ϵ-shrinking" the convex cells of this partition. In the terminology of Pivato (2013a), rules like this are called *balance rules*.[6] However, mean partition rules do not necessarily involve convex partitions, as the next examples show.

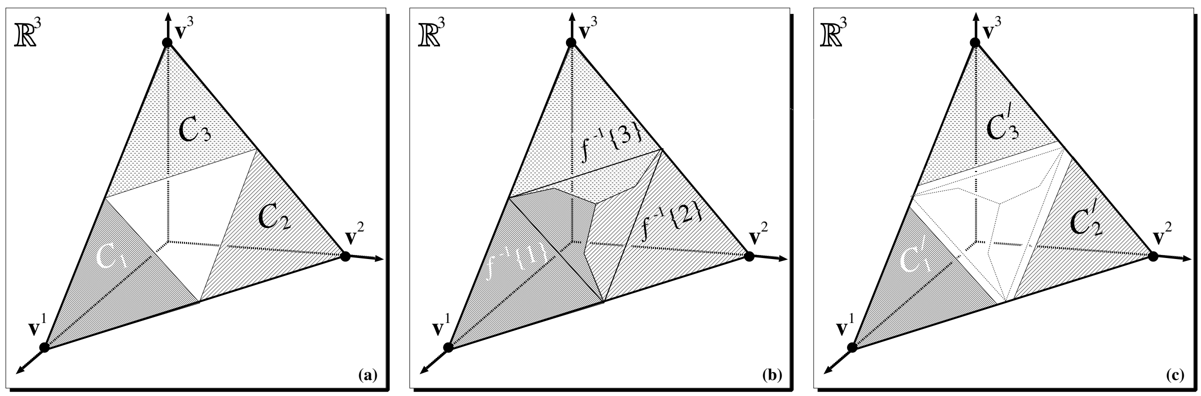

Figure 5: (Example 3.3(a)) Any majoritarian rule is a mean partition rule.

Example 3.3. (a) (*Majoritarian rules*) Let $\mathcal{S} = \{1, 2, \ldots, N\}$, let $\mathbb{V} := \mathbb{R}^N$, and define $\mathcal{V} := \{\mathbf{v}^1, \ldots, \mathbf{v}^N\}$ as in Example 3.1(b). Let $\mathcal{C}$ be the convex hull of $\mathcal{V}$, and for all $s \in \mathcal{S}$, let $\mathcal{C}_s := \{\mathbf{c} \in \mathcal{C};\ c_s > \frac{1}{2}\}$, as shown in Figure 5(a) Let $f : \mathcal{C} \longrightarrow \mathcal{S}$ be any function such that $\mathcal{C}_s \subseteq f^{-1}\{s\}$ for all $s \in \mathcal{S}$. Thus, if we define the rule F as in formula (1), then $F(\mathbf{V}) = s$ whenever more than half of all the voters support s; in other words, F is a ***majoritarian*** rule. If no alternative receives a clear majority, then the decision is determined by the structure of f in the part of $\mathcal{C}$ not covered by $\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_N$; in particular, note that the set $f^{-1}\{s\}$ need not be convex for any $s \in \mathcal{S}$. Figure 5(b) shows one possible example. To see that any majoritarian rule is a mean partition rule, let $\epsilon > 0$, and let $\mathcal{C}' := \mathcal{C}'_1 \sqcup \cdots \sqcup \mathcal{C}'_N$, where for all $s \in \mathcal{S}$, we define $\mathcal{C}'_s := \{\mathbf{c} \in \mathcal{C};\ c_s > \frac{1}{2} + \epsilon\}$, as shown in Figure 5(c).

(b) (*Condorcet consistent rules*) Let $\mathcal{S}$ be a finite set of alternatives, let $\mathcal{N}$ be a set containing exactly one of the pairs (s, t) or (t, s), for each $s, t \in \mathcal{S}$, and let $\mathbb{V} := \mathbb{R}^{\mathcal{N}}$. For any $\mathbf{v} \in \mathbb{V}$, we define $v_{s \succ t} := v_{s,t}$ if $(s, t) \in \mathcal{N}$, whereas $v_{s \succ t} := 1 - v_{t,s}$ if $(t, s) \in \mathcal{N}$. For any strict preference order $\succ$ on $\mathcal{S}$, let $\mathbf{v}^{\succ} \in \mathbb{V}$ be the unique vector such that $v^{\succ}_{s \succ t} = 1$ for all $s, t \in \mathcal{S}$. (In other words, for all $(s, t) \in \mathcal{N}$, we have $v^{\succ}_{s,t} := 1$ if $s \succ t$, whereas $v^{\succ}_{s,t} := 0$ if $t \succ s$.) Let $\mathcal{V} := \{\mathbf{v}^{\succ};\ \succ$ is a strict preference order on $\mathcal{S}\}$. This is the basis for a voting rule where each voter expresses a strict preference order over $\mathcal{S}$, and we keep track of the total support for each pairwise comparison. Let $\mathcal{C}$ be the convex hull of $\mathcal{V}$; see Figure 6(i) for the case $\mathcal{S} = \{a, b, c\}$. For each $s \in \mathcal{S}$, let $\mathcal{C}_s := \{\mathbf{c} \in \mathcal{C};\ c_{s \succ t} > \frac{1}{2}$ for all $t \neq s\}$, as shown in Figures 6(ii-iv). Let $f : \mathcal{C} \longrightarrow \mathcal{S}$ be a function such that $\mathcal{C}_s \subseteq f^{-1}\{s\}$ for all

[6] See Pivato (2013a) for the precise definition of balance rules and their axiomatic characterization.

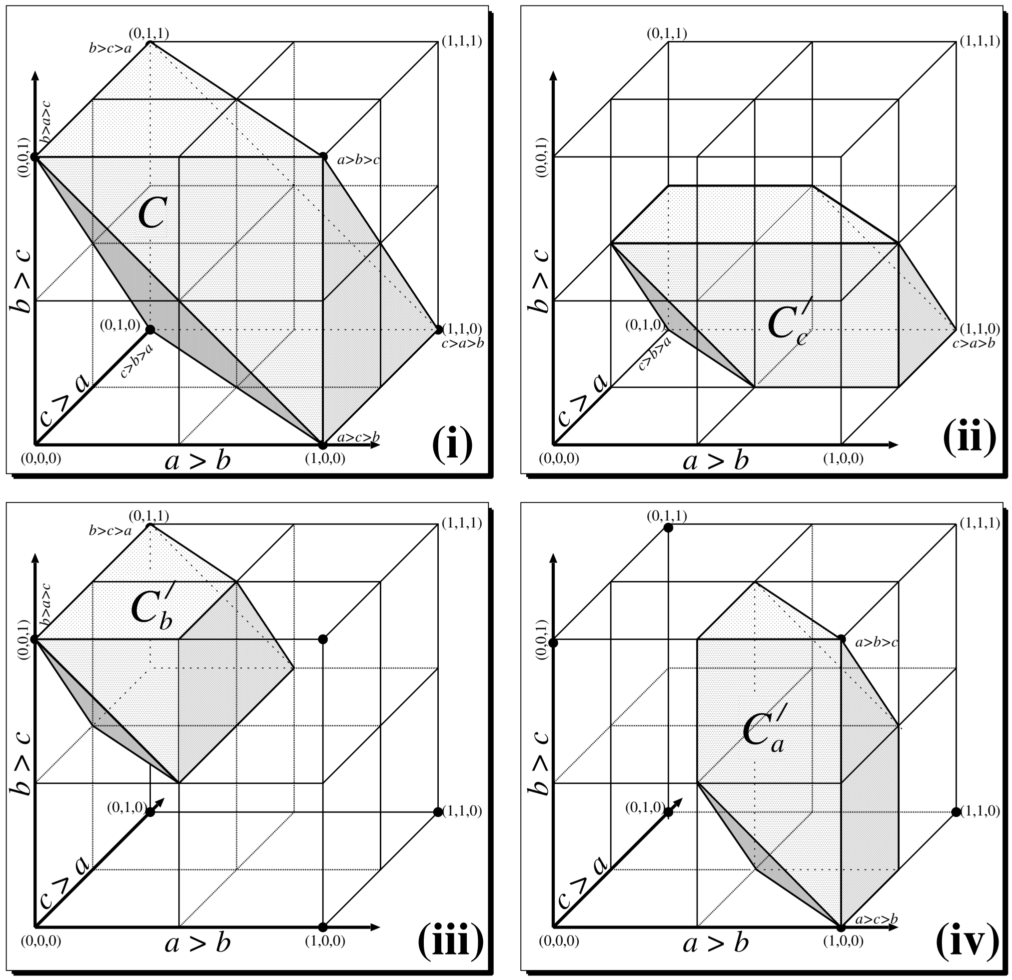

Figure 6: (Example 3.3(b)) Any Condorcet-consistent rule is a mean partition rule.

$s \in \mathcal{S}$. Thus, if we define F as in formula (1), then $F(\mathbf{V}) = s$ whenever s is the *Condorcet winner*, meaning that more than half of all the voters prefer s to each other alternative in $\mathcal{S}$. In other words, F is a *Condorcet-consistent* rule. If no alternative is a Condorcet winner, then the decision is determined by the structure of f in the part of $\mathcal{C}$ not covered by $\bigsqcup_{s\in\mathcal{S}} \mathcal{C}_s$. Many popular voting rules are Condorcet consistent, including Copeland rule, the Simpson-Kramer ("minimax") rule, the Tideman ("ranked pairs") rule, and the Schulze rule. For most of these rules, the set $f^{-1}\{s\}$ is not convex for any $s \in \mathcal{S}$. To see that any Condorcet-consistent rule is a mean partition rule, let $\epsilon > 0$, and let $\mathcal{C}' := \bigsqcup_{s\in\mathcal{S}} \mathcal{C}'_s$, where for all $s \in \mathcal{S}$, we define $\mathcal{C}'_s := \{\mathbf{c} \in \mathcal{C};\ \ c_{s\succ t} > \frac{1}{2} + \epsilon \text{ for all } t \neq s\}$. $\diamondsuit$

As Example 3.3 shows, the set $\mathcal{C}'$ which appears in (M3) and (M4) could actually be a rather small subset of $\mathcal{C}$. However, the smaller $\mathcal{C}'$ becomes, the more difficult it will be to satisfy the *Identification* condition we will introduce in Sections 4 and 5. In contrast, if a mean partition rule is based on a convex partition, then $\mathcal{C}'$ can be a very large subset of $\mathcal{C}$. Thus, while we do not *require* mean partition rules to use convex partitions, the *Identification* condition of Sections 4 and 5 is more easily satisfied for such rules.

All of the previous examples have assumed that $\mathcal{S}$ is *finite*. But there are also mean partition rules where $\mathcal{S}$ is infinite, or even a continuum, as shown by the next example.

Example 3.4. (*The average rule*) Let $\mathbb{V}$ be an inner product space (e.g. $\mathbb{V} = \mathbb{R}^N$), and let $\mathcal{S}$ be a convex subset of $\mathbb{V}$. Let $\mathcal{C} = \mathcal{V} = \mathcal{S}$, and let $f_{\text{ave}} : \mathcal{C} \longrightarrow \mathcal{S}$ be the identity function. This represents the rule where each voter declares an "ideal point" in $\mathcal{S}$, and the outcome is the arithmetic average of these ideal points. Note that (M2) and (M3) are satisfied (with $\mathcal{C}' := \mathcal{S}$ and δ arbitrary), because the identity function is uniformly continuous, and the preimage of each point is a singleton. $\diamondsuit$

4 Epistemic social choice with independent voters

The main focus of this paper is *correlated* voters. But for ease of reading, we will first introduce the main ideas in an environment with *independent* voters. Each voter is represented using a *behaviour model*: a function that maps each possible state of the world to a probability distribution over votes. A wide variety of behaviour models are mathematically possible, but most of these will not occur in an actual electorate of human voters, whose behaviour presumably conforms to certain psychological regularities and/or cultural norms. We will not explicitly model these psychological and cultural factors; instead, we will represent them implicitly by singling out a subset of possible behaviour models we call a *populace*. We will suppose that any actual electorate is constructed by sampling from this populace. The results of this section (Propositions 4.1 and 4.3) show that, if the populace satisfies certain conditions, then the mean partition rule applied to a large electorate of independent voters is highly likely to get the correct answer. By applying these results to some of the mean partition rules introduced in Examples 3.1 and 3.3, we rederive the most general versions of the Condorcet Jury Theorem which have appeared in previous literature (Example 4.2). We are also obtain a very general version of the Wisdom of Crowds principle (Example 4.4) and a CJT-type result for *log-likelihood scoring rules*, a class of voting rules which play a prominent role in "maximum-likelihood estimator" approaches to epistemic social choice theory (Example 4.5).

Let $\mathcal{S}$ be the metric space of the possible states of the world (the true state being unknown). Let $(\mathbb{V}, \mathcal{V}, F)$ be a mean partition rule taking outcomes in $\mathcal{S}$. Let $\mathcal{I}$ be a finite set of individuals, and let $I := |\mathcal{I}|$. We suppose that each individual's vote is a random variable, which is dependent on the true state of nature. The idea is that each individual obtains some information about the state of nature (possibly incomplete and/or incorrect), combines it with her own pre-existing beliefs, and formulates a belief about the state of nature, which she expresses using her vote. Our goal is to use the pattern of these votes to estimate the true state of nature.

Formally, for each individual $i \in \mathcal{I}$, we posit a *behaviour model* $\rho^i : \mathcal{S} \longrightarrow \Delta(\mathcal{V})$; if the true state is $s \in \mathcal{S}$, then the probability distribution of individual i's vote will be $\rho^i(s)$. For any $\mathbf{v} \in \mathcal{V}$, we will write $\rho^i(\mathbf{v}|s)$ for the value of $\rho^i(s)$ evaluated at $\mathbf{v}$ —i.e. the probability that individual i votes for $\mathbf{v}$, given that the true state is s. Let $\mathbb{E}[\rho(s)]$ denotes the *expected* *value* of a $\rho(s)$-random variable —in other words, the mean value of the distribution $\rho(s)$. If $\mathcal{C}$ is the closed convex hull of $\mathcal{V}$, then $\mathbb{E}[\rho(s)] \in \mathcal{C}$.[7]

Different voters may have different behaviour models (due to differing competency, different prior beliefs, or access to different information sources). Furthermore, it is not realistic to suppose that we have precise knowledge of the behaviour model of every voter (or even of *any* voter); in general, we only know some broad qualitative properties of their behaviour models. Thus, we will suppose that there is some set $\mathcal{P}$ of possible behaviour models (i.e. functions from $\mathcal{S}$ into $\Delta(\mathcal{V})$), and all we know is that $\rho^i \in \mathcal{P}$ for all $i \in \mathcal{I}$. We will refer to $\mathcal{P}$ as a ***populace*** on $\mathcal{V}$. Let $F = (\mathbb{V}, \mathcal{V}, f)$ be a mean partition rule, and let $\mathcal{C}$ be the closed convex hull of $\mathcal{V}$. We will say that a populace $\mathcal{P}$ is ***sagacious*** for F if there is some set $\mathcal{C}' \subseteq \mathcal{C}$ satisfying conditions (M3) and (M4) such that $\mathcal{P}$ satisfies two conditions:

Identification. For any $\rho \in \mathcal{P}$ and any $s \in \mathcal{S}$, the *expected* value of a $\rho(s)$-random variable lies in the f-preimage of s inside $\mathcal{C}'$. That is: $\mathbb{E}[\rho(s)] \in \mathcal{C}'$ and $f(\mathbb{E}[\rho(s)]) = s$.

Minimal Determinacy. There is some $M \geq 0$ such that $\mathrm{var}[\rho(s)] \leq M$ for all $\rho \in \mathcal{P}$ and $s \in \mathcal{S}$.

The *Identification* condition says that, while an individual's *actual* vote may be incorrect, the *expected value* of her vote indicates the true state of nature —at least once it has been "interpreted" using the function f. The variance of an individual's vote distribution is a measure of " randomness": if the variance is large, then this person's vote is quite unpredictable, and likely to be far from its expected value. *Minimal Determinacy* places a limit on the randomness of each voter.

Note that the epistemic *reliability* of a voter is determined both by the mean *and* the variance of her behaviour model —if $\rho^i(s)$ has a small variance, but its expected value is very close to the boundary of $f^{-1}\{s\}$, while $\rho^j(s)$ has a larger variance, but its expected value is much farther from the boundary of $f^{-1}\{s\}$, then it may turn out that voter j's opinion is a more *reliable* indicator of the true state of nature than voter i, even though voter j's opinion is also more *random*. It is for this reason that we use the term "determinacy" rather than "reliability" to describe the bound on variance.[8] Also note that, if the set $\mathcal{V}$ is bounded (in particular, if $\mathcal{V}$ is finite), then *Minimal Determinacy* is automatically satisfied (because there will be some M such that $\mathrm{var}(\rho) \leq M$ for any $\rho \in \Delta(\mathcal{V})$).

Our first result concerns the case when $\mathcal{S}$ is finite. It says that if a large number of voters are drawn from a sagacious populace, and their votes are independent random variables, then the output of the voting rule will be the true state of nature, with very high probability.

Proposition 4.1 *Let F be a mean partition rule ranging over a finite set $\mathcal{S}$, and let $\mathcal{P}$ be a populace which is sagacious for F. For all $i \in \mathbb{N}$, let $\rho_i \in \mathcal{P}$. Fix $s \in \mathcal{S}$, and*

[7] Formally, $\mathbb{E}[\rho(s)] := \int_{\mathcal{V}} \mathbf{v} \; \mathrm{d}\rho(s)[\mathbf{v}]$. This $\mathbb{V}$-valued integral is defined by taking a limit in $\mathbb{V}$; this is why we defined $\mathcal{C}$ to be the *closed* convex hull of $\mathcal{V}$ in (M2), and not merely its convex hull. If $\mathbb{V}$ is finite-dimensional, then this integral is defined in the obvious way. But if $\mathbb{V}$ is infinite-dimensional, then it is a Bochner integral; for details, see Remark A.3 in the Appendix.

[8] I thank the referee for emphasizing the importance of this distinction.

suppose $\{\mathbf{v}_i\}_{i=1}^{\infty}$ *are all independent random variables, where, for all* $i \in \mathbb{N}$, $\mathbf{v}_i$ *is drawn from distribution* $\rho_i(s)$. *Then* $\lim_{I\to\infty} \text{Prob}\left[F\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_I\right) = s\right] = 1.$

Example 4.2. (a) (*Condorcet Jury Theorem*) Let $\mathcal{S} = \mathcal{V} := \{\pm 1\}$ and let F_{maj} be as in Example 3.1(a). Let $\mathcal{P}$ be the set of all behaviour models $\rho : \{\pm 1\} \longrightarrow \Delta\{\pm 1\}$ such that $\rho(s|s) > \frac{1}{2} + \epsilon$ (and thus, $\rho(-s|s) < \frac{1}{2} - \epsilon$) for both $s \in \{\pm 1\}$. Let $\mathcal{C}'_{-1} := [-1, -\epsilon)$ and $\mathcal{C}'_1 := (\epsilon, 1]$. Then $\mathbb{E}[\rho(s)] \in \mathcal{C}'_s$ for any $\rho \in \mathcal{P}$ and $s \in \{\pm 1\}$. Thus, $F_{\text{maj}}\left(\mathbb{E}[\rho(s)]\right) = s$, so *Identification* is satisfied. Furthermore, $\text{var}(\rho) < 4$ for any $\rho \in \Delta\{\pm 1\}$, so *Minimal Determinacy* is always satisfied. Thus, Proposition 4.1 yields an extension of the Condorcet Jury Theorem to heterogenous voters, originally stated by Paroush (1998): If the voter's opinions about some dichotomous choice are independent random variables, and each voter satisfies some minimal level of competency (i.e. her probability of identifying the correct answer is ϵ-better than a coin flip), then the outcome of a simple majority vote will converge in probability to the correct answer as the voting population becomes large.

(b) (*Plurality CJT*) Let $N \geq 2$, and let $\mathcal{S} := \{1, 2, \ldots N\}$. Define $(\mathbb{V}, \mathcal{V}, F_{\text{plu}})$ as in Example 3.1(b). Let $\mathcal{P}$ be the set of all behaviour models $\rho : \mathcal{S} \longrightarrow \Delta(\mathcal{V})$ such that $\rho(\mathbf{v}^s|s) > \rho(\mathbf{v}^t|s) + \epsilon$, for all $s, t \in \mathcal{S}$ with $s \neq t$. For all $s \in \mathcal{S}$, define $\mathcal{C}_s$ as in Example 3.1(b). Then $\mathbb{E}[\rho(s)] = (\rho^i(1|s), \rho^i(2|s), \ldots, \rho^i(N|s)) \in \mathcal{C}_s$ for all $\rho \in \mathcal{P}$ and $s \in \mathcal{S}$; thus, *Identification* is satisfied. Furthermore, $\text{var}(\rho) < N$ for any $\rho \in \Delta(\mathcal{V})$, so *Minimal Determinacy* is always satisfied. Thus, Proposition 4.1 yields a "polychotomous" extension of the CJT, originally stated by Goodin and List (2001; Proposition 2): if each voter has some minimal level of competency (i.e. is ϵ-better than a random guess), then the outcome of the plurality rule will converge in probability to the correct answer as the voting population becomes large.

By applying a similar argument to Examples 3.3(a,b), we could also develop polychotomous versions of the CJT for majoritarian and Condorcet-consistent voting rules. $\diamondsuit$

In fact, Proposition 4.1 is a special case of the next result, which also applies when $\mathcal{S}$ is infinite. This result says that, if a large number of voters are drawn from a sagacious populace, and their votes are independent random variables, then the output of the voting rule will be *very close* to the true state of nature, with very high probability.

Proposition 4.3 *Let* F *be a mean partition rule ranging over an arbitrary set* $\mathcal{S}$, *and let* $\mathcal{P}$ *be a populace which is sagacious for* F. *For all* $i \in \mathbb{N}$, *let* $\rho_i \in \mathcal{P}$. *Fix* $s \in \mathcal{S}$, *and suppose* $\{\mathbf{v}_i\}_{i=1}^{\infty}$ *are all independent random variables, where, for all* $i \in \mathbb{N}$, $\mathbf{v}_i$ *is drawn from distribution* $\rho_i(s)$. *Then for any open subset* $\mathcal{U} \subset \mathcal{S}$ *containing* s, *we have* $\lim_{I\to\infty} \text{Prob}\left[F\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_I\right) \in \mathcal{U}\right] = 1.$

Example 4.4. (*The Wisdom of Crowds*) Let $\mathbb{V}$ be an inner product space (e.g. $\mathbb{V} = \mathbb{R}^N$), let $\mathcal{V} = \mathcal{S}$ be a convex subset of $\mathbb{V}$, and let F_{ave} be the average rule, as in Example 3.4. Fix $M > 0$, and let $\mathcal{P}$ be the set of all behaviour models $\rho : \mathcal{S} \longrightarrow \Delta(\mathcal{V})$ such that, for all $\mathbf{s} \in \mathcal{S}$, $\mathbb{E}[\rho(\mathbf{s})] = \mathbf{s}$ and $\text{var}[\rho(s)] \leq M$. Then *Identification* and *Minimal Determinacy* are satisfied. Thus, Proposition 4.3 yields the Wisdom of Crowds principle for the estimation of some real-valued (or, more generally, vector-valued) quantity: if each voter estimates the quantity, and their estimates are independent, unbiased, and have finite variance, then the average of their estimates will converge in probability to the correct answer. $\diamondsuit$

The classic examples of the Wisdom of Crowds involve a numerical quantity (e.g. the weight of an ox). But Example 4.4 also applies when $\mathbb{V}$ is a vector space —even an infinite-dimensional vector space. For example, let $\mathbb{V}$ be the space of all continuous real-valued functions on an interval $[a, b]$, equipped with the inner product $\langle v, w\rangle := \int_a^b v(r)\, w(r)\, \mathrm{d}r$ for all $v, w \in \mathbb{V}$. Many decision problems involve estimating such functions. For example, an oligopolistic firm must estimate the shape of the demand curve to determine its optimal pricing strategy. A central bank must estimate the functional relationship between the Consumer Price Index and other macroeconomic variables, to determine whether it should intervene in the money supply. And the IPCC must estimate the functional relationship between atmospheric CO_2 levels and other meteorological variables. Each expert might have her own opinion, and the committee must aggregate these opinions to obtain a group estimate. Example 4.4 says that, under certain conditions, a large enough committee can obtain a good estimate by averaging the opinions of the committee members.

Another interesting application is *probabilistic opinion pooling* (Genest and Zidek, 1986; Clemen and Winkler, 1999). Let $\mathcal{X}$ be a finite set, and let $\Delta(\mathcal{X})$ be the probability simplex in $\mathbb{R}^{\mathcal{X}}$. We interpret $\mathcal{X}$ as the space of possible resolutions of some uncertainty (e.g. the weather or the stock market next Tuesday). Each voter has an opinion about this uncertainty, in the form of a probability vector in $\Delta(\mathcal{X})$. We wish to aggregate these opinions to construct the best "collective opinion" in $\mathcal{S}$. If we define $\mathbb{V} := \mathbb{R}^{\mathcal{X}}$ and $\mathcal{V} := \mathcal{S} := \Delta(\mathcal{X})$, then the average rule of Example 4.4 is called the *linear pooling rule*: the collective opinion is the *average* of the opinions of the voters.

There is another approach to probabilistic opinion pooling. Let $\Delta_+(\mathcal{X})$ denote the set of probability vectors with full support on $\mathcal{X}$. For any $\mathbf{p} = (p_x)_{x\in\mathcal{X}}$ in $\Delta_+(\mathcal{X})$, let $\log(\mathbf{p}) := [\log(p_x)]_{x\in\mathcal{X}}$, an element of $\mathbb{R}^{\mathcal{X}}$. Let $\mathcal{V}_{\log} := \{\log(\mathbf{p});\ \ \mathbf{p} \in \Delta_+(\mathcal{X})\}$, and let $\mathcal{C}_{\log}$ be the closed convex hull of $\mathcal{V}_{\log}$ in $\mathbb{R}^{\mathcal{X}}$. Define $f_{\log} : \mathcal{C}_{\log} \longrightarrow \Delta_+(\mathcal{X})$ as follows: for any $\mathbf{c} = (c_x)_{x\in\mathcal{X}}$ in $\mathcal{C}_{\log}$, we define $f_{\log}(\mathbf{c}) := (e^{c_x}/K)_{x\in\mathcal{X}}$, where $K := \sum_{x\in\mathcal{X}} e^{c_x}$. The resulting mean partition rule $F_{\log} = (\mathbb{R}^{\mathcal{X}}, \mathcal{V}_{\log}, f_{\log})$ is called the *logarithmic pooling rule*.[9] In effect, this rule takes the *geometric* average of the opinions of the individual voters, and renormalizes it to obtain a probability vector.

Proposition 4.3 can be invoked to obtain Wisdom of Crowds justifications for both the linear and logarithmic pooling rules, by specifying a suitable populace $\mathcal{P}$. In the interests of brevity, we will suppress the details. The next example shows an entirely different way that logarithmic probabilities can arise.

Example 4.5. (*Log-likelihood scoring rules*) Let $\mathcal{S}$ be a finite set. Let $p : \mathcal{S} \longrightarrow \Delta(\mathcal{S})$ be a function (called an *error model*). For any $s, t \in \mathcal{S}$, we interpret $p(t|s)$ be the probability that a voter will *believe* that the true state is t, when it is actually s. Let $\mathbb{V} := \mathbb{R}^{\mathcal{S}}$, and for all $r \in \mathcal{S}$, define $\mathbf{v}^r := (v_s^r)_{s\in\mathcal{S}} \in \mathbb{V}$ by setting $v_s^r := \log[p(r|s)]$, for all $s \in \mathcal{S}$. Let

[9] $F_{\log}$ is a mean partition rule because $f_{\log}$ is uniformly continuous on $\mathcal{C}_{\log}$. To see this, note that $f_{\log}$ is differentiable, and for any $\mathbf{c} \in \mathcal{C}_{\log}$, if $f_{\log}(\mathbf{c}) = \mathbf{p}$, then $\partial_x f_{\log}(\mathbf{c})_x = p_x - p_x^2$ for all $x \in \mathcal{X}$, while $\partial_y f_{\log}(\mathbf{c})_x = -p_x\, p_y$ for all $x \neq y \in \mathcal{X}$. Thus, $|\partial_y f_{\log}(\mathbf{c})_x| < 1$ for all $x, y \in \mathcal{X}$; uniform continuity follows.

$\mathcal{V} := \{\mathbf{v}^r\}_{r\in\mathcal{S}}$, let $\mathcal{C}$ be the convex hull of $\mathcal{V}$, and let $f^p_{\log} := f_{\rm scr} : \mathcal{C}{\longrightarrow}\mathcal{S}$ be the scoring rule defined in Example 3.1(f). We will refer to this as a *log-likelihood* scoring rule.

Assume the votes of the different voters are independent random variables (conditional on the true state of nature). Any error model p' induces a behaviour model ρ' by setting $\rho'(\mathbf{v}^r|s) := p'(r|s)$ for all $r, s \in \mathcal{S}$. For any $\eta > 0$, let $\mathcal{P}_{p,\eta}$ be the populace consisting of all behaviour models ρ' induced by an error model p' such that $|p'(t|s) - p(t|s)| < \eta$ for all $t, s \in \mathcal{S}$. If $p(t|s) > 0$ for all $t, s \in \mathcal{S}$, then the populace $\mathcal{P}_{p,\eta}$ satisfies *Minimal Determinacy* (see Proposition A.2(a) in the Appendix). Now fix $\epsilon > 0$, and for all $s \in \mathcal{S}$, define $\mathcal{C}^\epsilon_s := \{\mathbf{c} \in \mathcal{C};\ \ c_s > c_t + \epsilon \text{ for all } t \neq s\}$. If $\mathcal{C}'_\epsilon := \bigcup_{s\in\mathcal{S}} \mathcal{C}^\epsilon_s$, then $f^p_{\log}$ satisfies (M3) when restricted to $\mathcal{C}'$. If ϵ and η are small enough, then $\mathcal{P}_{p,\eta}$ satisfies *Identification* with respect to $f^p_{\log}$ and $\mathcal{C}'_\epsilon$ (see Proposition A.2(b) in the Appendix).

Thus, Proposition 4.1 yields an extension of the Condorcet Jury Theorem to *any* log-likelihood scoring rule. If a sufficiently large number of independent random voters are drawn from the populace $\mathcal{P}_{p,\eta}$, then the log-likelihood scoring rule $F^p_{\log}$ will select the true state of nature, with probability arbitrarily close to 1. For example, if $\mathcal{S}$ is the space of preference orders on some set of alternatives, then this conclusion holds for the Kemeny rule, given the error model proposed by Young (1986, 1988, 1995, 1997). $\diamondsuit$

For any error model p, the outcome of the rule $F^p_{\log}$ defined in Example 4.5 will be the maximum likelihood estimator (MLE) of the true state.[10] Conversely, *any* scoring rule can be interpreted as a log-likelihood scoring rule for some error model, and in many cases, these are in fact maximum likelihood estimators (Pivato, 2013b, Theorem 2.2(a,b)). For example, the Kemeny rule (Example 3.1(e)) is the MLE for a natural error model on the space of preference orders (Young, 1986, 1988, 1995, 1997). More generally, on any metric space $(\mathcal{S}, d)$ which is "sufficiently symmetric", the generalized median rule (Example 3.1(d)) is the MLE for any *exponential* error model, where $p(s|t) = C\,\exp[-\alpha\, d(s,t)]$, for some constants $\alpha, C > 0$ (Pivato, 2013b, Corollary 3.2).[11] For example, the median rule has been proposed as an MLE for equivalence relations and other binary relations (Régnier, 1977; Barthélémy and Monjardet, 1981, 1988). Example 4.5 is a complementary result: not only is $F^p_{\log}$ an MLE, but it is highly likely to identify the true state, in the large-population limit. (For a similar result, see Xia (2015, Theorem 1 and Example 3).)

Examples 4.2 and 4.4 are well-known results from epistemic social choice theory. But Example 4.5 is new, as is the next and last example of this section.

Example 4.6. (*The wisdom of the median voter*) Let $\mathcal{S}$ be a finite subset of $\mathbb{R}$, and let F be the median voting rule from Example 3.1(c). Let $\epsilon > 0$, and let $\mathcal{P}$ be the set of all behaviour models $\rho : \mathcal{S}{\longrightarrow}\Delta(\mathcal{V})$ such that, for all $s \in \mathcal{S}$, $\sum_{r<s} \rho_s(\mathbf{v}^r) < \frac{1}{2} - \epsilon$ and $\sum_{r\leq s} \rho_s(\mathbf{v}^r) > \frac{1}{2} + \epsilon$ (and hence, $\rho_s(\mathbf{v}^s) > 2\epsilon$). If $\mathcal{C}'_s$ is the set defined Example 3.1(d), then it is easily verified that this error model satisfies *Identification*. Meanwhile, *Minimal*

[10] This follows from Theorem 2.2(b) of Pivato (2013b). Note that $F^p_{\log}$ is "balanced" scoring rule (in the terminology of Pivato (2013b)) because the way $f^p_{\log}$ is defined from the error model p.

[11] In particular, this is the case if $(\mathcal{S}, d)$ has a transitive group of isometries. For example, a sphere has this degree of symmetry. But in fact, a weaker (but more technical) condition is sufficient.

Determinacy is automatically satisfied because $\mathcal{V}$ is finite. Thus, Proposition 4.3 says that the *median estimate* of a large group of voters will be close to the correct value, with high probability. $\diamondsuit$

The error model in Example 4.6 may seem somewhat unrealistic, since each voter must have a positive probability of exactly identifying the correct value. But ϵ could be extremely small, so this is not as restrictive as it seems. Also, the *median error* of each voter must be zero. This would be plausible if we had reason to believe that the error distribution of each voter was symmetric around zero (e.g. a normal distribution). But it might be implausible in other scenarios.

5 Correlated Voters

The problem with the model in Section 4 is its assumption that the errors of the voters are stochastically independent. We will now extend this model to allow for correlated errors. To model such correlations, we introduce a *collective behaviour model*: a function that maps each possible state of the world to a probability distribution over profiles. A wide variety of collective behaviour models are mathematically possible, but most of these will not occur in reality because the collective behaviour of an actual electorate is a partly determined by sociological, political and economic factors, the educational system and the communications infrastructure, among other things. We will not explicitly model these factors; instead, we will represent them implicitly by focusing on a subset of possible collective behaviour models, which we call a *culture*. We will suppose that any actual electorate is drawn from this culture. In particular, any populace from Section 4 yields such a culture (see Example 5.1). For any culture, we define two functions, σ and κ; the former measures the indeterminacy of the average voter, while the latter measures the correlation between voters. The main results of this section (Proposition 5.2 and Theorem 5.3) say that, if σ is constant and κ decays as the population grows, then the mean partition rule applied to a large electorate will get the correct answer with very high probability.

Culture. If the voters are correlated, then we can no longer consider their vote distributions separately. Instead, we must consider the *joint* distribution of all the voters. Given a set $\mathcal{I}$ of individuals and a set $\mathcal{V}$ of votes, a ***profile*** is an element $\mathbf{V} = (\mathbf{v}_i)_{i\in\mathcal{I}}$ of $\mathcal{V}^{\mathcal{I}}$, which assigns a vote $\mathbf{v}_i$ to each individual i in $\mathcal{I}$. A ***collective behaviour model*** on $\mathcal{V}$ is a function $\rho : \mathcal{S}\longrightarrow\Delta(\mathcal{V}^{\mathcal{I}})$, which determines a probability distribution $\rho(s)$ over the set of possible profiles, for each possible state $s \in \mathcal{S}$. We cannot assume that we have detailed knowledge of the collective behaviour model of a society. We will only suppose that it arises from some family of collective behaviour models with certain statistical properties. For this reason, we define a ***culture*** on $\mathcal{V}$ to be a sequence $\mathfrak{R} = (\mathcal{R}_I)_{I=1}^{\infty}$ where, for all $I \in \mathbb{N}$, $\mathcal{R}_I$ is a set of collective behaviour models on $\mathcal{V}$, for a population of size I. Note that a culture is not intended as a description of a *single* society facing a single epistemic problem. It describes an infinite family of *possible* societies, of all possible sizes, facing a family of possible decision problems.

Correlation. We will need to quantify the correlation between voters arising from a culture. Let $I \in \mathbb{N}$ and let $\mathcal{I} := [1 \ldots I]$. Fix a collective behaviour model $\rho : \mathcal{S} \longrightarrow \Delta(\mathcal{V}^{\mathcal{I}})$, and some state $s \in \mathcal{S}$. For all $i \in \mathcal{I}$, let

$$\widehat{\mathbf{v}}_i \quad := \quad \int_{\mathcal{V}} \mathbf{v}_i \ \mathrm{d}\rho[\mathbf{V}|s]$$

be the expected value of individual i's vote, given the state s.

Let $\langle \_, \_ \rangle$ be an inner product on $\mathbb{V}$. Fix $s \in \mathcal{S}$, and let $\mathbf{V} = (\mathbf{v}_i)_{i \in \mathcal{I}}$ be a $\rho(s)$-random profile. For any $i \in \mathcal{I}$, the random vector $(\mathbf{v}_i - \widehat{\mathbf{v}}_i)$ measures the amount by which individual i's vote deviates from its expected value (if the voters satisfy *Identification*, then we can think of this as the "error" in i's vote). The inner product $\langle \mathbf{v}_i - \widehat{\mathbf{v}}_i,\ \mathbf{v}_j - \widehat{\mathbf{v}}_j \rangle$ measures the extent to which the errors of voters i and j are "aligned" with respect to the geometry of $\mathbb{V}$. The ***covariance*** of voters i and j is the expected value of this inner product:

$$\mathrm{cov}(\mathbf{v}_i, \mathbf{v}_j) \quad := \quad \mathbb{E}[\langle \mathbf{v}_i - \widehat{\mathbf{v}}_i,\ \mathbf{v}_j - \widehat{\mathbf{v}}_j \rangle].$$

This measures the amount, *on average*, by which we can expect the errors of i and j to align in same direction in $\mathbb{V}$. Note that $\mathrm{var}[\mathbf{v}_i] = \mathrm{cov}(\mathbf{v}_i, \mathbf{v}_i)$. We then define the ***covariance matrix*** of $\rho(s)$ to be the $I \times I$ matrix $\mathrm{cov}[\rho(s)] := [b_{i,j}]_{i,j=1}^{I}$, where, for all $i, j \in [1 \ldots I]$, $b_{i,j} := \mathrm{cov}(\mathbf{v}_i, \mathbf{v}_j)$.

It is important to note that $b_{i,j}$ measures the covariance of *errors*, not the covariance of *votes*. For example, suppose that i and j were not only perfectly reliable, but that there was some $\mathbf{v} \in \mathcal{V}$ with $F(\mathbf{v}) = s$ such that $\mathbf{v}_i = \mathbf{v}_j = \mathbf{v}$ with probability 1. Then their *votes* would be perfectly correlated, but we would have $b_{i,j} = 0$, since their *error* terms would both be zero. Likewise, if $b_{i,j} < 0$, this means that the *errors* of i and j are anticorrelated —it does not mean their *votes* are anticorrelated.

Since we do not know the true collective behaviour model of society, and we don't know the true state of nature, we also do not know the true covariance matrix of the voters. We can only assume that it comes from some family satisfying certain broad qualitative properties. For this reason, we define a ***covariance structure*** to be a sequence $\mathfrak{B} = (\mathcal{B}_I)_{I=1}^{\infty}$, where, for all $I \in \mathbb{N}$, $\mathcal{B}_I$ is a collection of $I \times I$ symmetric, positive semidefinite matrices. The elements of $\mathcal{B}_I$ are the *possible* covariance matrices that we could see in a society of size I. We say that $\mathfrak{B}$ is the covariance structure of the culture $\mathfrak{R}$ if, for every $I \in \mathbb{N}$, $\mathcal{B}_I$ is the set of all covariance matrices $\mathrm{cov}[\rho(s)]$, for any $\rho \in \mathcal{R}_I$ and $s \in \mathcal{S}$.

For any collective behaviour model $\rho \in \mathcal{R}_I$, and any state $s \in \mathcal{S}$, the covariance matrix $\mathbf{B} = \mathrm{cov}[\rho(s)]$ combines two sorts of information: the diagonal entries encode the " randomness" of individual voters, whereas the off-diagonal entries encode the correlations *between* voters. To be precise, for any $i \in [1 \ldots I]$, the diagonal entry $b_{i,i}$ is the variance of individual i's vote in a $\rho(s)$-random profile. For any distinct $i, j \in [1 \ldots I]$, the off-diagonal entry $b_{i,j}$ is the covariance between the error of individual i's vote and the error of individual j's vote, in a $\rho(s)$-random profile. (Note that $b_{i,j}$ could be negative, reflecting *anti*correlation between the errors of i and j.) For this reason, we will associate two distinct numerical values with each covariance matrix $\mathbf{B} \in \mathcal{B}_I$. We define

$$\sigma(\mathbf{B}) \;:=\; \frac{1}{I}\sum_{i=1}^{I} b_{i,i}, \qquad \text{and} \qquad \kappa(\mathbf{B}) \;:=\; \frac{1}{I(I-1)}\sum_{\substack{i,j=1\\ i\neq j}}^{I} b_{i,j}. \tag{2}$$

In other words, $\sigma(\mathbf{B})$ is the average of the diagonal entries (i.e., the average variance of the voters' errors), while $\kappa(\mathbf{B})$ is the average of the off-diagonal entries (i.e., the average covariance *between* the voters' errors).

Let $F = (\mathbb{V}, \mathcal{V}, f)$ be a mean partition rule, and let $\mathcal{C}$ be the closed convex hull of $\mathcal{V}$. Let $\mathfrak{R} = (\mathcal{R}_I)_{I=1}^{\infty}$ be a culture on $\mathcal{V}$, with covariance structure $(\mathcal{B}_I)_{I=1}^{\infty}$. We will say that $\mathfrak{R}$ is *sagacious* with respect to F if there exists some set $\mathcal{C}' \subseteq \mathcal{C}$ satisfying conditions (M3) and (M4), such that $\mathfrak{R}$ satisfies the following three properties relative to $\mathcal{C}'$ and $\langle \_, \_ \rangle$.

Identification. For any $I \in \mathbb{N}$, any $\rho \in \mathcal{R}_I$, and any $s \in \mathcal{S}$, if $(\mathbf{v}_i)_{i\in\mathcal{I}}$ is a $\rho(s)$-random profile, then for all $i \in [1 \ldots I]$, the *expected* value of $\mathbf{v}_i$ is in the f-preimage of s inside $\mathcal{C}'$ —i.e. $\mathbb{E}_{\rho(s)}[\mathbf{v}_i] \in f^{-1}\{s\} \cap \mathcal{C}'$.

Asymptotic Determinacy. For any $I \in \mathbb{N}$, let $\sigma(I) := \sup_{\mathbf{B}\in\mathcal{B}_I} \sigma(\mathbf{B})$. Then $\lim_{I\to\infty} \frac{\sigma(I)}{I} = 0$.

Asymptotically Weak Average Covariance. For any $I \in \mathbb{N}$, let $\kappa(I) := \sup_{\mathbf{B}\in\mathcal{B}_I} \kappa(\mathbf{B})$. Then $\lim_{I\to\infty} \kappa(I) = 0$.

Here, the key condition is *Asymptotically Weak Average Covariance*. This says that voters' errors can be correlated, but as the society grows large, the *average* covariance between the errors of different voters must become small. *Identification* has exactly the same interpretation as in Section 4. The condition of *Asymptotic Determinacy* is a very weak form of the *Minimal Determinacy* condition from Section 4. To see, this, first note that *Minimal Determinacy* could be weakened to the following condition:

Average Determinacy. There is some constant $M > 0$ such that, for any $I \in \mathbb{N}$, and any $\mathbf{B} \in \mathcal{B}_I$, $\sigma(\mathbf{B}) < M$.

This condition allows some voters to be *very* unpredictable, as long as the *average* variance of the voters is bounded.[12] Clearly, *Minimal Determinacy* implies *Average Determinacy*. But *Asymptotic Determinacy* is even weaker than *Average Determinacy*: it says that even the *average* variance of the voters can grow with population size, as long as it does not grow too quickly. (To be precise: its growth rate must be sublinear.)

Example 5.1. Let F be a mean partition rule, and let $\mathcal{P}$ be a sagacious populace for F, as defined in Section 4. Given any behaviour models $\rho_1, \ldots, \rho_I \in \mathcal{P}$, and any $s \in \mathcal{S}$, let $\rho_1 \otimes \cdots \otimes \rho_I(s)$ be the product probability measure on $\mathcal{V}^I$ —that is, the distribution of a

[12] For a version of the CJT assuming *Average Determinacy*, see Grofman (1989, Theorem II). For a version of the CJT with a condition similar to Asymptotic Determinacy, see Boland (1989, Theorem 3).

random profile where $\mathbf{v}_1, \ldots, \mathbf{v}_I$ are are independent random variables, with $\mathbf{v}_i$ distributed according to $\rho_i(s)$ for all $i \in [1 \ldots I]$. This yields a collective behaviour model $\rho_1 \otimes \cdots \otimes \rho_I : \mathcal{S} \longrightarrow \Delta(\mathcal{V}^I)$. For all $I \in \mathbb{N}$, define $\mathcal{R}_I := \{\rho_1 \otimes \cdots \otimes \rho_I;\ \ \rho_1, \ldots, \rho_I \in \mathcal{P}\}$, and then let $\mathfrak{R} := (\mathcal{R}_I)_{I=1}^{\infty}$. Then $\mathfrak{R}$ is a sagacious culture for F. $\diamondsuit$

The next result says that, if $\mathcal{S}$ is finite, and a random profile of votes is drawn from a sagacious culture, and the population is sufficiently large, then with very high probability, the outcome of the voting rule will be the true state of nature.

Proposition 5.2 *Let F be a mean partition rule ranging over a finite set $\mathcal{S}$, and let $(\mathcal{R}_I)_{I=1}^{\infty}$ be a sagacious culture for F. For all $I \in \mathbb{N}$, let $\rho_I \in \mathcal{R}_I$. Then for any $s \in \mathcal{S}$,*

$$\mathrm{Prob}\left(F(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_I) = s \ \middle|\ (\mathbf{v}_i)_{i=1}^{I} \text{ is a } \rho_I\text{-random profile} \right) \xrightarrow[I \to \infty]{} 1.$$

In fact, Proposition 5.2 is a special case of our main result, which also applies when $\mathcal{S}$ is infinite. It says that, if a random profile of votes is drawn from a sagacious culture, and the population is sufficiently large, then with very high probability, the outcome of the voting rule will be *very close* to the true state of nature.

Theorem 5.3 *Let F be a mean partition rule ranging over a set $\mathcal{S}$, and let $(\mathcal{R}_I)_{I=1}^{\infty}$ be a sagacious culture for F. For all $I \in \mathbb{N}$, let $\rho_I \in \mathcal{R}_I$. Then for any $s \in \mathcal{S}$, and any open set $\mathcal{U} \subset \mathcal{S}$ containing s,*

$$\mathrm{Prob}\left(F(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_I) \in \mathcal{U} \ \middle|\ (\mathbf{v}_i)_{i=1}^{I} \text{ is a } \rho_I\text{-random profile} \right) \xrightarrow[I \to \infty]{} 1.$$

In the case of dichotomous choice (i.e. the classical Condorcet Jury Theorem), this result is very similar to a result proved by Ladha (1992)[13]. Theorem 5.3 extends this result to a much larger family of epistemic social choice rules, and also weakens *Minimal Determinacy* to *Asymptotic Determinacy*. Using Theorem 5.3, it is straightforward to extend Examples 4.2, 4.4, 4.5 and 4.6 to a setting with correlated voters; we leave the details to the reader.

To obtain a sagacious culture —in particular, to satisfy *Asymptotically Weak Average Correlation* —we need to make $\kappa(I)$ small. To this end, we could try to reduce the positive correlation between voters —i.e. reduce the number and magnitude of positive entries in the covariance matrices in $\mathfrak{B}$. But we could also increase the *anti*correlation between voters —i.e. increase the number and magnitude of *negative* entries in these covariance matrices. One could cultivate such anticorrelation by maximizing the *cognitive diversity* of the voter population (Page 2008, Chapter 8; Landemore 2013, Section 6.3).[14] One could also maximize the diversity of information and opposing opinions to which voters are exposed; this is a strong argument for maximal freedom of the press in democratic polities

[13]See Ladha (1992, Corollary, p.628) and Ladha (1995, Proposition 1).

[14]Another argument for cognitive diversity treats collective decisions as creative problem-solving processes, akin to massively multidimensional nonlinear optimization problems (Page, 2008; Landemore, 2013). But this is totally unrelated to the "anticorrelation" argument presented here.

(Ladha, 1992). It is also the basis for the adversarial legal system favoured in common-law jury trials. Finally, such anticorrelation can arise when voters split into opposing factions or political parties, as occurs in parliamentary debates.[15] Indeed, Theorem 5.3 remains true if we replace *Asymptotically Weak Average Correlation* with the weaker condition that $\limsup_{I\to\infty} \kappa(I) \leq 0$ —in particular, $\kappa(I)$ could be negative. Thus, for epistemic democracy, there is no such thing as "too much" anticorrelation between voters.

In general, a culture might be sagacious with respect to some voting rules, and not sagacious with respect to others. But in some cases, a culture can be sagacious in a way that is *independent* of the choice of voting rule. Let $\mathcal{S}$ be a finite set, let $\mathbb{V}$ be an inner product space, and let $\mathfrak{R} = (\mathcal{R}_I)_{I=1}^{\infty}$ be a culture on $\mathbb{V}$. We will say that $\mathfrak{R}$ is ***identifiable*** if, for all $s \in \mathcal{S}$, there is a compact, convex subset $\mathcal{K}_s \subset \mathbb{V}$ such that, for any $I \in \mathbb{N}$ and $\rho \in \mathcal{R}_I$, if $(\mathbf{v}_i)_{i\in\mathcal{I}}$ is a $\rho(s)$-random profile, then for all $i \in [1 \ldots I]$, the *expected* value of $\mathbf{v}_i$ is in $\mathcal{K}_s$. We also require $\mathcal{K}_s$ and $\mathcal{K}_t$ to be disjoint for any distinct $s, t \in \mathcal{S}$; this is a minimal condition for any possibility of epistemic social choice. We will say that a covariance structure $\mathfrak{B} = (\mathcal{B}_I)_{I=1}^{\infty}$ is ***sagacious*** if it satisfies *Asymptotic Determinacy* and *Asymptotically Weak Average Covariance.*

Proposition 5.4 *If $\mathfrak{R}$ is an identifiable culture, with covariance structure $\mathfrak{B}$, and $\mathfrak{B}$ is sagacious, then there is a mean partition rule F on $\mathbb{V}$ such that $\mathfrak{R}$ is sagacious for F.*

So, what sort of covariance structures are sagacious? We now turn to this question.

6 Social networks

This section explores covariance structures arising from *social networks*; our goal is derive sagacity of the covariance structure from the geometry of the network. We will not work with a *specific* social network, but rather, with an entire family of social networks, of all possible sizes —we call this a *social web*. We first consider a scenario where each voter is only correlated with her nearest neighbours in the network. In this case, the resulting covariance structure will be sagacious as long as the average voter does not acquire new neighbours "too quickly" as the population increases (Proposition 6.2). In particular, this result applies to social networks with power law degree distributions, which arise frequently in applications (Example 6.1). We then consider a more general model, where voters can be correlated even if they are not neighbours. In this case, there is a tradeoff between two asymptotics: the asymptotic *decay rate* of the covariance between voters as a function of their distance in the social network, and the asymptotic *growth rate* of the "sphere of radius r" around a typical voter, as r becomes large —roughly speaking, this measures the "dimension" of the network. In this case, the resulting covariance structure will be sagacious as long as the correlations decay quickly enough to balance the sphere-growth rate (Proposition 6.5). In particular, for a finite-dimensional network, it is sufficient for the voters to have an *exponential* covariance decay rate (Example 6.4(a)).

[15]However, the epistemic deficiencies due to correlation *within* each faction might outweigh the epistemic benefits of anticorrelation *between* factions.

Social webs. A *graph* is a set $\mathcal{I}$ equipped with a symmetric, reflexive binary relation $\sim$. If $\mathcal{I}$ is a set of voters, then we can interpret a graph as a *social network*: if $i \sim j$, we interpret this to mean that voters i and j are somehow "socially connected" (e.g. friends, family, neighbours, colleagues, classmates, etc.).

We cannot assume that we have *exact* knowledge of the social network topology; we can only assume that belongs to some family of graphs satisfying broad qualitative properties. For this reason, we define a *social web* to be a sequence $\mathfrak{N} = (\mathcal{N}_I)_{I=1}^{\infty}$, where, for all $I \in \mathbb{N}$, $\mathcal{N}_I$ is a set of possible graphs of size I. Thus, our hypotheses will be formulated in terms of the asymptotic properties of the graphs in $\mathcal{N}_I$, as $I \to \infty$. But before we can formulate these hypotheses, we need some basic concepts from graph theory.

Sublinear average degree growth. For any $i \in \mathcal{I}$, the *degree* of i is the number of links i has in the graph $(\mathcal{I}, \sim)$. Formally, $\deg(i, \sim) := \#\{j \in \mathcal{I};\ \ i \sim j\}$. If $|\mathcal{I}| = I$, then the *average degree* of the graph $(\mathcal{I}, \sim)$ is defined:

$$\operatorname{avedeg}(\mathcal{I}, \sim) \quad := \quad \frac{1}{I} \sum_{i \in \mathcal{I}} \deg(i, \sim).$$

This is the average number of social links of a voter in the social network described by $(\mathcal{I}, \sim)$. We then define $\overline{\operatorname{avedeg}}(\mathcal{N}_I) := \sup_{(\mathcal{I},\sim)\in\mathcal{N}_I} \operatorname{avedeg}(\mathcal{I}, \sim)$. We will say that a social web $(\mathcal{N}_I)_{I=1}^{\infty}$ exhibits *sublinear average degree growth* if

$$\lim_{I\to\infty} \frac{1}{I} \overline{\operatorname{avedeg}}(\mathcal{N}_I) \quad = \quad 0. \tag{3}$$

For instance, if $\overline{\operatorname{avedeg}}(\mathcal{N}_I)$ remains bounded as $I \to \infty$, then the limit (3) is obviously satisfied. However, the limit (3) even allows $\overline{\operatorname{avedeg}}(\mathcal{N}_I)$ to grow as $I \to \infty$, as long as it grows more slowly than a linear function.

Example 6.1. (*Asymptotic degree distributions*) Let $(\mathcal{I}, \sim)$ be a graph. For all $n \in \mathbb{N}$, let

$$\mu_{(\mathcal{I},\sim)}(n) \quad := \quad \frac{1}{I}\ \#\{i \in \mathcal{I}\ ;\ \deg(i, \sim) = n\}.$$

This defines a probability distribution $\mu_{(\mathcal{I},\sim)} \in \Delta(\mathbb{N})$, called the the *degree distribution* of $(\mathcal{I}, \sim)$. If $\mu \in \Delta(\mathbb{N})$ is another probability distribution, then we define the distance between μ and $\mu_{(\mathcal{I},\sim)}$ by

$$d(\mu, \mu_{(\mathcal{I},\sim)}) \quad := \quad \sum_{n=1}^{\infty} n \cdot \left|\mu_{(\mathcal{I},\sim)}(n) - \mu(n)\right|.$$

We will say that a social web $\mathfrak{N}$ has *asymptotic degree distribution* μ if

$$\lim_{I\to\infty} \sup_{(\mathcal{I},\sim)\in\mathcal{N}_I} d(\mu, \mu_{(\mathcal{I},\sim)}) \quad = \quad 0.$$

Let $\text{avedeg}(\mu) := \sum_{n=1}^{\infty} \mu(n)\, n$. If this value is finite, and $\mathfrak{N}$ has asymptotic degree distribution μ, then it is easy to check that $\overline{\text{avedeg}}(\mathcal{N}_I)$ will converge to $\text{avedeg}(\mu)$ as $I\to\infty$; thus, $\mathfrak{N}$ will have sublinear average degree growth.

For example, many social networks seem to exhibit a "power law" degree distribution of the form $\mu(n) \approx K/n^{\alpha}$, for all $n \in \mathbb{N}$, where $\alpha > 1$, and where $K > 0$ is a normalization constant (Barabási and Albert, 1999; Albert et al., 1999). This is a well-defined probability distribution on $\mathbb{N}$, as long as $\alpha > 1$. (Typically, $2 < \alpha < 3$.) Networks with power law distributions often contain a surprisingly large number of "superconnected" or "hub" individuals, whose degrees are much larger than that of the typical person. Thus, in such networks, some individuals can be correlated with a very large number of other individuals. However, $\text{avedeg}(\mu)$ is still finite, as long as $\alpha > 2$. Thus, if a social web has a power law asymptotic degree distribution with $\alpha > 2$, then it will have sublinear average degree growth. $\diamondsuit$

Not all social webs have sublinear average degree growth. For example, if $\alpha < 2$ in Example 6.1, then $\overline{\text{avedeg}}(\mathcal{N}_I)$ will grow at a superlinear rate as $I\to\infty$. For another example, suppose $\mathcal{N}_I$ is generated by sampling the Erdös-Renyi "random graph" model, where there is a constant probability p that any two randomly chosen agents are linked. Then $\overline{\text{avedeg}}(\mathcal{N}_I) \approx p\, I$, which grows linearly as $I\to\infty$. However, these are not considered realistic models for social networks in most situations, because the (Poisson) asymptotic degree distribution of the Erdös-Renyi model is a poor fit to the empirical data (Albert et al., 1999; Newman et al., 2002).

Nearest-neighbour covariance structures. Let $\mathfrak{B} = (\mathcal{B}_I)_{I=1}^{\infty}$ be a covariance structure, and let $\mathfrak{N} = (\mathcal{N}_I)_{I=1}^{\infty}$ be a social web. We will say that $\mathfrak{B}$ is a ***nearest-neighbour covariance structure*** for $\mathfrak{N}$ if:

- For any $I \in \mathbb{N}$ and $\mathbf{B} \in \mathcal{B}_I$, there is some graph $(\mathcal{I}, \sim)$ in $\mathcal{N}_I$ and some identification of $\mathcal{I}$ with $[1 \ldots I]$ such that, for all $i, j \in [1 \ldots I]$, we have $b_{i,j} \neq 0$ only if $i \sim j$,
- There is some constant $M > 0$ such that, for any $I \in \mathbb{N}$ and $\mathbf{B} \in \mathcal{B}_I$, we have $|b_{i,j}| \leq M$ for all $i, j \in [1 \ldots I]$.

We now come to the first result of this section.

Proposition 6.2 *If a social web $\mathfrak{N}$ has sublinear average degree growth, then any nearest-neighbour covariance structure for $\mathfrak{N}$ is sagacious.*

In fact, Proposition 6.2 is only a special case of the main result of this section. But before we can state this result, we need more terminology.

Generalized degrees. Let $(\mathcal{I}, \sim)$ be a connected graph. A ***path*** in $(\mathcal{I}, \sim)$ is a sequence of vertices $i_0, i_1, \ldots, i_L \in \mathcal{I}$ such that $i_0 \sim i_1 \sim \cdots \sim i_L$; we say this path has ***length*** L, and that it ***connects*** i_0 to i_L. For any $i, j \in \mathcal{I}$, let $d_\sim(i,j)$ be the length of the shortest path connecting i to j in $(\mathcal{I}, \sim)$. For completeness, we also define $d_\sim(i,i) := 0$ for all $i \in \mathcal{I}$. Observe that $d_\sim$ is a metric on $\mathcal{I}$. (It is called the ***geodesic metric*** of the graph.) For any $r \in \mathbb{N}$ and $i \in \mathcal{I}$, we define the ***r-degree*** of i as $\deg^r(i, \sim) := \#\{j \in \mathcal{I};\ d_\sim(i,j) = r\}$. Thus, $\deg^1(i, \sim)$ is just the degree of i, as defined above. Now let $\gamma : \mathbb{N} \longrightarrow [0, \infty]$ be a function (typically, increasing). For any $i \in \mathcal{I}$, we define the ***γ-degree*** of i by

$$\deg^\gamma(i, \sim) \quad := \quad \sup_{r \in \mathbb{N}} \frac{\deg^r(i, \sim)}{\gamma(r)}. \tag{4}$$

We then define

$$\begin{aligned} \operatorname{avedeg}^\gamma(\mathcal{I}, \sim) \quad &:= \quad \frac{1}{I} \sum_{i \in \mathcal{I}} \deg^\gamma(i, \sim), & (5) \\ \text{and} \qquad \overline{\operatorname{avedeg}}^\gamma(\mathcal{N}_I) \quad &:= \quad \sup_{(\mathcal{I}, \sim) \in \mathcal{N}_I} \operatorname{avedeg}^\gamma(\mathcal{I}, \sim). & (6) \end{aligned}$$

We will say that a social web $(\mathcal{N}_I)_{I=1}^\infty$ exhibits ***sublinear average γ-degree growth*** if

$$\lim_{I \to \infty} \frac{1}{I} \overline{\operatorname{avedeg}}^\gamma(\mathcal{N}_I) \quad = \quad 0. \tag{7}$$

For instance, suppose we define $\gamma_1 : \mathbb{N} \longrightarrow \{1, \infty\}$ by

$$\gamma_1(r) \quad := \quad \begin{cases} 1 & \text{if} \quad r = 1; \\ \infty & \text{if} \quad r \geq 2. \end{cases} \tag{8}$$

Then clearly, $\deg^{\gamma_1}(i, \sim) = \deg(i, \sim)$ for all $i \in \mathcal{I}$ and all $(\mathcal{I}, \sim) \in \mathcal{N}_I$. Thus, formula (7) is equivalent to formula (3); thus, a social web will have sublinear average γ_1-degree growth if and only if it has sublinear average degree growth.

Example 6.3. (*Social networks from infinite graphs*) Let $\mathcal{J}$ be an infinite set of vertices, and let $\sim$ be a graph structure on $\mathcal{J}$; this is called an ***infinite graph***. If $\gamma : \mathbb{N} \longrightarrow [0, \infty]$ is some function, then $(\mathcal{J}, \sim)$ has ***γ-bounded growth*** if we have $\deg^r(j, \sim) \leq \gamma(r)$, for all $j \in \mathcal{I}$ and all $r \in \mathbb{N}$. In other words, $\deg^\gamma(j) \leq 1$ for all $j \in \mathcal{J}$.

For example, if $(\mathcal{J}, \sim)$ is the infinite two-dimensional grid shown in Figure 7(a), then $\deg^r(i) = 4\,r$ for all $r \in \mathbb{N}$; thus, $(\mathcal{J}, \sim)$ has growth bounded by the function $\gamma(r) := 4\,r$. More generally, if $(\mathcal{J}, \sim)$ is an infinite subgraph of a two-dimensonal grid, like the one shown in Figure 7(b), then its growth bounded by the function $\gamma(r) := 4\,r$. Likewise, if $(\mathcal{J}, \sim)$ was an infinite subgraph of a D-dimensional grid, then it would have growth bounded by a polynomial function $\gamma(r) := K\,r^{D-1}$ (for some constant $K > 0$). As these examples show, a graph with a "D-dimensional" geometry has polynomially bounded growth of degree $D - 1$.

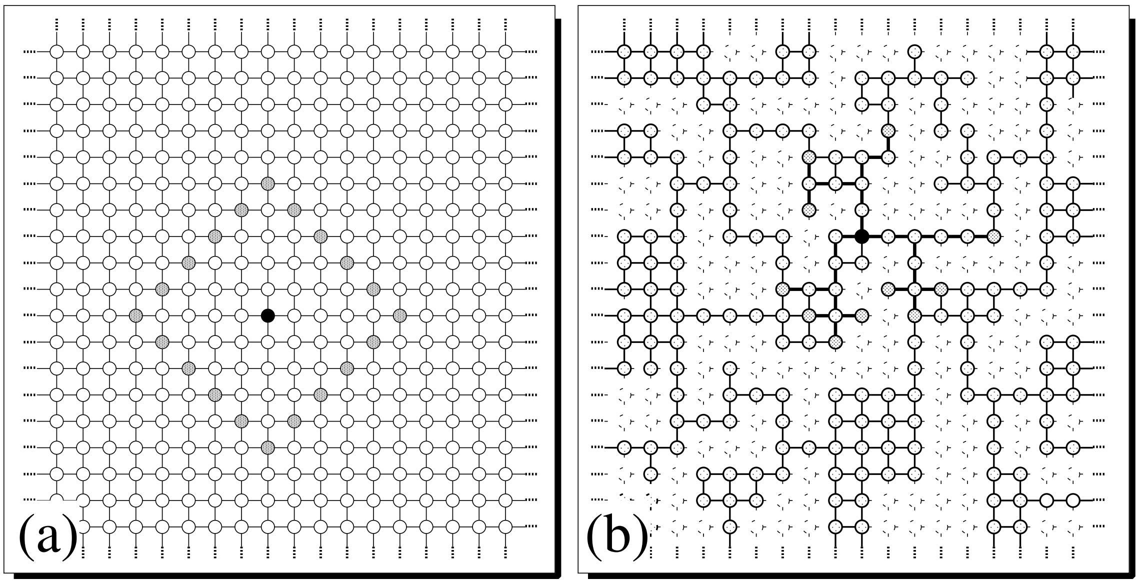

Figure 7: (Example 6.3) **(a)** An infinite, two-dimensional grid has growth bounded by $\gamma(r) = 4\,r$. For example, if i is the black node, then $\deg^5(i, \sim) = 20$ (the number of grey nodes). **(b)** If $(\mathcal{J}, \sim)$ is an infinite subgraph of a two-dimensional grid, then its growth is also bounded by $\gamma(r) = 4\,r$. In this case, if i is the black node, then $\deg^5(i, \sim) = 9$.

In contrast, suppose $(\mathcal{J}, \sim)$ is an infinite tree where every node has degree 3, as shown in Figure 8(a). Then $(\mathcal{J}, \sim)$ has growth bounded by $\gamma(r) = 3\,(2^{r-1})$. More generally, if $M \in \mathbb{N}$, and $(\mathcal{J}, \sim)$ is any graph where every vertex has degree $(M+1)$ or less, then $(\mathcal{J}, \sim)$ has growth bounded by the exponential function $\gamma(r) := M^r$.

For all $I \in \mathbb{N}$, let $\mathcal{N}_I$ be a collection of connected subgraphs of $(\mathcal{J}, \sim)$ with exactly I vertices; then the sequence $\mathfrak{N} = (\mathcal{N}_I)_{I=1}^{\infty}$ is a social web, which we will say is ***subordinate*** to $(\mathcal{J}, \sim)$. Heuristically, the vertices in the graph $(\mathcal{J}, \sim)$ represent the set of all "potential" people who could exist, and the links in $(\mathcal{J}, \sim)$ are all "potential" social connections between them. Thus, any *actual* social network will be some finite subgraph of $(\mathcal{J}, \sim)$; these are the graphs which appear in $\mathfrak{N}$. If $(\mathcal{J}, \sim)$ has growth bounded by the function γ, then it is easy to see that $\overline{\mathrm{avedeg}}^{\gamma}(\mathcal{N}_I) \leq 1$ for all $I \in \mathbb{N}$; thus, the asymptotic condition (7) is trivially satisfied, so that $\mathfrak{N}$ has sublinear average γ-degree growth. $\diamondsuit$

Correlation decay. Let $(\mathcal{I}, \sim)$ be a graph, and let $\mathbf{B} \in \mathbb{R}^{I \times I}$ be an $I \times I$ matrix (e.g. a covariance matrix). Let $\beta : \mathbb{N} \longrightarrow \mathbb{R}_+$ be a function (typically, decreasing). We will say that the matrix $\mathbf{B}$ exhibits ***β-decay*** relative to $(\mathcal{I}, \sim)$ if (after bijectively identifying $\mathcal{I}$ with $[1 \ldots I]$ in some way), we have $b_{i,j} \leq \beta[d_\sim(i,j)]$ for all $i, j \in \mathcal{I}$. In particular, $\mathbf{B}$ exhibits ***exponential decay*** if there are some constants $\lambda \in (0,1)$ and $K \geq 0$ such that $b_{i,j} \leq K\,\lambda^{d_\sim(i,j)}$ for all $i, j \in \mathcal{I}$. Exponential correlation decay is a typical phenomenon in the spatially distributed stochastic processes studied in statistical physics, such as Ising models of ferromagnetism (Penrose and Lebowitz, 1974; Procacci and Scoppola, 2001; Bach

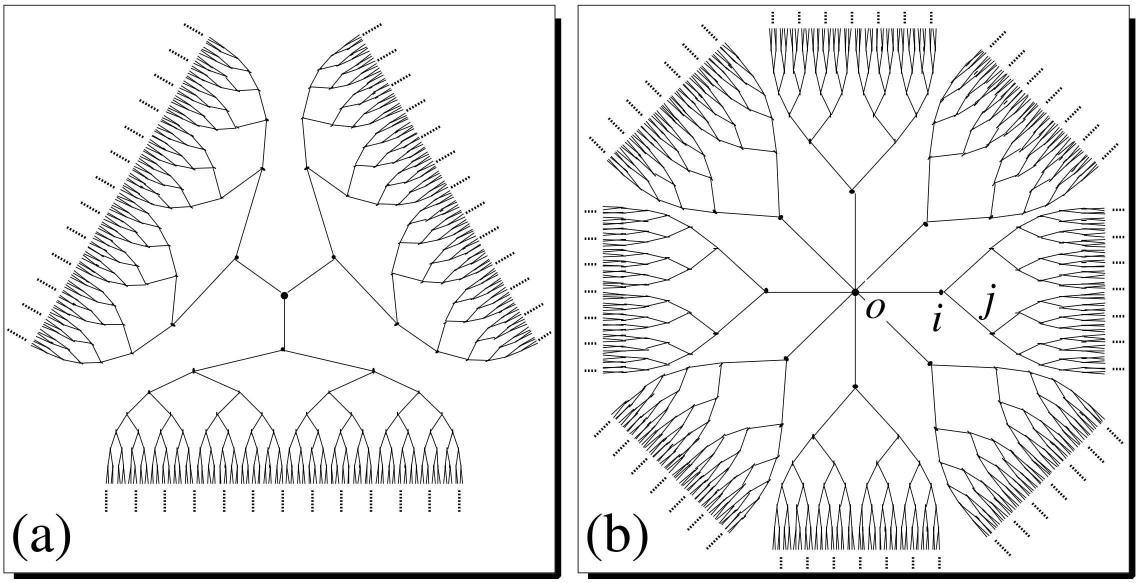

Figure 8: (Example 6.3) **(a)** If $(\mathcal{J}, \sim)$ is an infinite tree where all nodes have 3 edges, then its growth is bounded by $\gamma(r) = 3\,(2^{r-1})$. **(b)** If $(\mathcal{J}, \sim)$ is eight infinite binary trees around a hub, then its growth is bounded by $\gamma(r) = 8\,(2^{r-1})$.

and Møller, 2003). The opinions of the voters in a social network can be seen as such a spatially distributed stochastic process.

We will say that a covariance structure $\mathfrak{B} = (\mathcal{B}_I)_{I=1}^{\infty}$ exhibits β- *covariance decay* relative to social web $\mathfrak{N} = (\mathcal{N}_I)_{I=1}^{\infty}$ if, for every $I \in \mathbb{N}$, and every matrix $\mathbf{B} \in \mathcal{B}_I$, there is some graph $(\mathcal{I}, \sim)$ in $\mathcal{N}_I$ such that $\mathbf{B}$ exhibits β-decay relative to $(\mathcal{I}, \sim)$. For example, let $M > 0$, and define $\beta(1) := M$ while $\beta(r) := 0$ for all $r \geq 2$. Then $\mathfrak{B}$ exhibits β- covariance decay relative to $\mathfrak{N}$ if and only if $\mathfrak{B}$ is a nearest-neighbour covariance structure for $\mathfrak{N}$.

Subordinate covariance structures. We will say that a covariance structure $\mathfrak{B}$ is *subordinate* to a social web $\mathfrak{N}$ if there exist functions $\beta : \mathbb{N} \longrightarrow \mathbb{R}^+$ and $\gamma : \mathbb{N} \longrightarrow [0, \infty]$ such that $\mathfrak{N}$ has sublinear average γ-degree growth, $\mathfrak{B}$ exhibits β- covariance decay relative to $\mathfrak{N}$, and also

$$\sum_{r=0}^{\infty} \gamma(r)\,\beta(r) \quad < \quad \infty. \tag{9}$$

(Here, we adopt the convention that $\infty \cdot 0 = 0$.) Note that the faster $\gamma(r)$ grows as $r \to \infty$, the faster β must decay to zero in order for inequality (9) to be satisfied.

Example 6.4. (a) Let $M, D \in \mathbb{N}$ and suppose that $\mathfrak{N}$ is subordinate to an infinite, D-dimensional grid or an M-ary tree, as described in Example 6.3. Let $\gamma(r) := M^r$ for all $r \in \mathbb{N}$; then $\mathfrak{N}$ has sublinear average γ-degree growth. Let $\lambda < 1/M$, let $\beta(r) := \lambda^r$ for all $r \in \mathbb{N}$; and suppose that every matrix in $\mathfrak{B}$ exhibits β-exponential covariance decay with respect to some graph in $\mathfrak{N}$. Let $c := M\,\lambda$; then $0 < c < 1$, and

$$\sum_{r=0}^{\infty} \gamma(r)\,\beta(r) \quad = \quad \sum_{r=0}^{\infty} M^r\,\lambda^r \quad = \quad \sum_{r=0}^{\infty} c^r \quad = \quad \frac{1}{1-c} \quad < \quad \infty.$$

Thus, inequality (9) is satisfied, so $\mathfrak{B}$ is subordinate to $\mathfrak{N}$.

(b) Suppose $\mathfrak{N}$ has sublinear average degree growth, and $\mathfrak{B}$ is a nearest-neighbour covariance structure for some social web $\mathfrak{N}$. As we have seen, this means there is some constant $M > 0$ such that $\beta(r) := M$ if $r = 1$ and $\beta(r) := 0$ for all $r > 0$, and $\mathfrak{B}$ exhibits β- covariance decay relative to $\mathfrak{N}$. Now define $\gamma_1 : \mathbb{N} \longrightarrow \{1, \infty\}$ by formula (8). Then inequality (9) is automatically satisfied. By comparing formulae (3) and (7), we see that $\mathfrak{N}$ has sublinear average γ_1-degree growth. Thus, $\mathfrak{B}$ is subordinate to $\mathfrak{N}$. $\diamondsuit$

We now come to the main result of this section.

Proposition 6.5 *Let $\mathfrak{N}$ be a social web. Then any covariance structure which is subordinate to $\mathfrak{N}$ is sagacious.*

For example, Proposition 6.2 follows by applying Proposition 6.5 to Example 6.4(b).

7 Deliberation

A growing literature argues that *deliberation* can improve the epistemic efficacy of democratic decision-making (Elster, 1998; Fishkin and Laslett, 2003; Landemore and Elster, 2012; Landemore, 2013). Deliberation can edify voters, so that they hold more informed, objective, and nuanced opinions. But it can also increases correlation between voters, perhaps leading to "groupthink". It is possible that the groupthink effect outweighs the edification effect, so that on the balance, deliberation leads to *worse* decisions. However, this section offers some evidence that this need not occur: we will show that, under certain hypotheses, the sagacity of a culture is preserved under a simple model of deliberation. This does not prove that deliberation makes groups *smarter* (our simple model ignores edification effects). But at least deliberation doesn't necessarily make groups stupider.

We will adapt a well-known model of deliberation proposed by DeGroot (1974):[16] we represent a deliberative institution as a family of linear transformations which can be applied to the profile of (vector-valued) opinions of the voters; in effect, these transformations replace each voter's opinion with a weighted average of her own opinion and those of her peers. We call such institutions *local* if no single voter has too strong an influence over other voters in this averaging processs. We show that local deliberative institutions cannot convert a sagacious culture into a non-sagacious culture (Proposition 7.1)

Let $\mathcal{I}$ be a set of voters. For all distinct $i, j \in \mathcal{I}$, let $d_{i,j} \geq 0$ be the "influence" of voter j on voter i. This could be determined by the level of respect or trust which i has

[16]For an interesting recent application of the DeGroot model, see Golub and Jackson (2010).

for j. Note that influence is not symmetric: we may have $d_{i,j} \neq d_{j,i}$. The diagonal entry $d_{i,i}$ measures i's confidence in her own opinions. Let $\mathbf{D} := [d_{ij}]_{i,j\in\mathcal{I}}$. We will assume that $\mathbf{D}$ is a stochastic matrix —that is, $\sum_{j\in\mathcal{I}} d_{i,j} = 1$, for all $i \in \mathcal{I}$. We will refer to $\mathbf{D}$ as an *influence matrix*. We cannot assume exact knowledge of the pattern of social influences in the society. Thus, instead of fixing a single influence matrix $\mathbf{D}$, we will consider an entire family of such influence matrices. Formally, we define a *deliberative institution* to be a sequence $\mathfrak{D} = (\mathcal{D}_I)_{I=1}^{\infty}$, where for all $I \in \mathbb{N}$, $\mathcal{D}_I$ is a family of $I \times I$ influence matrices.

A deliberative institution is not a culture. It is a *transformation*, which can be applied to a culture to obtain another culture, as we now explain. For the rest of this section, suppose that $\mathcal{V}$ is a *convex* subset of a vector space $\mathbb{V}$. Let $\mathbf{V} = (\mathbf{v}_i)_{i=1}^{I}$ be an I-voter profile in $\mathcal{V}^I$. Given an $I \times I$ stochastic matrix $\mathbf{D}$ (e.g. an element of $\mathcal{D}_I$), we define $\mathbf{D}\cdot\mathbf{V}$ to be the profile $\mathbf{V}' = (\mathbf{v}'_i)_{i=1}^{I}$, where, for all $i \in \mathcal{I}$,

$$\mathbf{v}'_i \quad := \quad \sum_{j=1}^{I} d_{i,j}\,\mathbf{v}_j.$$

For all $i \in \mathcal{I}$, $\mathbf{v}_i$ represents the opinion of voter i *before* deliberation, while $\mathbf{v}'_i$ represents her opinion *after* deliberation —it is a weighted average of her own opinion and those of her peers, with the weights reflecting their degree of "influence" over her.

Let $\rho : \mathcal{S} \longrightarrow \Delta(\mathcal{V}^{\mathcal{I}})$ be a collective behaviour model on $\mathcal{V}$, fix $s \in \mathcal{S}$, and suppose $\mathbf{V} = (\mathbf{v}_i)_{i=1}^{I}$ is a $\rho(s)$-random profile. Then $\mathbf{D}\cdot\mathbf{V}$ is another random profile. We denote the probability distribution of $\mathbf{D}\cdot\mathbf{V}$ by $\mathbf{D}\odot\rho(s)$. If we do this for all $s \in \mathcal{S}$, then we obtain a collective behaviour model $\mathbf{D}\odot\rho : \mathcal{S} \longrightarrow \Delta(\mathcal{V}^I)$.

Now, let $\mathfrak{R} = (\mathcal{R}_I)_{I=1}^{\infty}$ be a culture on $\mathcal{V}$, and let $\mathfrak{D} = (\mathcal{D}_I)_{I=1}^{\infty}$ be a deliberative institution. For all $I \in \mathbb{N}$, we define

$$\mathcal{D}_I \odot \mathcal{R}_I \quad := \quad \{\mathbf{D}\odot\rho\ ;\ \mathbf{D}\in\mathcal{D}_I \text{ and } \rho\in\mathcal{R}_I\}.$$

This is a collection of collective behaviour models on a population of I voters. Heuristically, it has the following interpretation:

- $\mathcal{R}_I$ is the set of collective behaviour models which could exist *before* deliberation.
- $\mathcal{D}_I$ is the set of the possible deliberations which could occur.
- $\mathcal{D}_I\odot\mathcal{R}_I$ is the set of the collective behaviour models which can exist *after* deliberation.

We then define the culture $\mathfrak{D}\odot\mathfrak{R} := (\mathcal{R}'_I)_{I=1}^{\infty}$, where, for each $I \in \mathbb{N}$, $\mathcal{R}'_I := \mathcal{D}_I \odot \mathcal{R}_I$. We interpret this as the culture which arises when voters drawn from the culture $\mathfrak{R}$ deliberate according to $\mathfrak{D}$.

For any $j \in \mathcal{I}$, we define $\overline{d}_j := \sum_{i\in\mathcal{I}} d_{i,j}$. This measures the "total influence" of voter j on other voters. A deliberative institution $\mathfrak{D}$ is *local* if there exists a constant $D > 0$ (which we will call the *modulus* of $\mathfrak{D}$) such that, for all $I \in \mathbb{N}$ and all $\mathbf{D} \in \mathcal{D}_I$ we have $\overline{d}_j \leq D$ for all $j \in \mathcal{I}$. In other words, the total influence of each voter in any society is bounded;

she can have a significant influence over at most a small number of individuals (although she might also have a very small influence over a much larger number of individuals). In particular, there are no "demagogues" who can strongly influence a large number of people.

Proposition 7.1 *Let $F = (\mathbb{V}, \mathcal{V}, f)$ be a mean partition voting rule, where $\mathcal{V}$ is a convex subset of $\mathbb{V}$. If $\mathfrak{D}$ is a local deliberative institution, and the culture $\mathfrak{R}$ is sagacious for F, then the culture $\mathfrak{D} \odot \mathfrak{R}$ is also sagacious for F.*

To illustrate the scope of this result, we will now construct some examples of local deliberative institutions. Given two deliberative institutions $\mathfrak{D}$ and $\mathfrak{E}$, we define $\mathfrak{D} \cdot \mathfrak{E} := (\mathcal{C}_I)_{I=1}^{\infty}$, where for all $I \in \mathbb{N}$, $\mathcal{C}_I := \{\mathbf{D}\,\mathbf{E};\ \mathbf{D} \in \mathcal{D}_I \text{ and } \mathbf{E} \in \mathcal{E}_I\}$. Informally, $\mathfrak{D} \cdot \mathfrak{E}$ represents a deliberative institution where the voters first deliberate according to an influence matrix drawn from $\mathfrak{E}$, and then deliberate further using a matrix drawn from $\mathfrak{D}$.

Given any $q \in [0,1]$, we define $q\,\mathfrak{D} + (1-q)\,\mathfrak{E} := (\mathcal{C}_I)_{I=1}^{\infty}$, where for all $I \in \mathbb{N}$, $\mathcal{C}_I := \{q\,\mathbf{D} + (1-q)\mathbf{E};\ \mathbf{D} \in \mathcal{D}_I \text{ and } \mathbf{E} \in \mathcal{E}_I\}$. Informally, this represents a deliberative institution where the influence of one voter on another is a weighted average of two forms of influence; one described by $\mathfrak{D}$ and the other by $\mathfrak{E}$. (For example, $\mathfrak{D}$ might describe influences arising from personal affection, while $\mathfrak{E}$ describes influences arising from professional respect and admiration.)

Proposition 7.2 *Let $\mathfrak{D}$ and $\mathfrak{E}$ be two local deliberative institutions. Then $\mathfrak{D} \cdot \mathfrak{E}$ is also local, and $q\,\mathfrak{D} + (1-q)\,\mathfrak{E}$ is local for any $q \in [0,1]$.*

For any deliberative institution $\mathfrak{D}$ and any $n \in \mathbb{N}$, we define $\mathfrak{D}^n := (\mathcal{D}_I^n)_{I=1}^{\infty}$, where for all $I \in \mathbb{N}$, $\mathcal{D}_I^n := \{\mathbf{D}_1 \cdots \mathbf{D}_n;\ \mathbf{D}_1, \ldots \mathbf{D}_n \in \mathcal{D}_I\}$. Informally, $\mathfrak{D}^n$ represents a deliberative institution where the voters deliberate n times, using n influence matrices drawn from $\mathfrak{D}$. Let $\mathfrak{D}^0 := \{\mathbf{I}\}$, where $\mathbf{I}$ is the identity matrix (this represents *no* deliberation). Finally, given any sequence $\mathbf{q} = (q_n)_{n=0}^{\infty}$ in $[0,1]$ with $\sum_{n=0}^{\infty} q_n = 1$, we can define the institution $\sum_{n=0}^{\infty} q_n \mathfrak{D}^n$ in the obvious way; informally, this is an institution where voters have deliberated a very large number of times, and the total influence of one voter on another is a weighted average of more direct, short-term effects (corresponding to small values of n) and more indirect, longer-terms effects (corresponding to larger values of n).

Corollary 7.3 *If $\mathfrak{D}$ is a local deliberative institution with modulus D, then $\sum_{n=0}^{\infty} q_n \mathfrak{D}^n$ is local as long as $\sum_{n=0}^{\infty} q_n\, D^n < \infty$.*

As a simple example, suppose $\mathcal{D}_I$ contains only one matrix, $\mathbf{D}$, and furthermore, suppose that most of the entries in $\mathbf{D}$ are zero. For any $i, j \in \mathcal{I}$, write "$j \rightsquigarrow i$" if $d_{i,j} > 0$. Informally, this means "j has some direct influence on i". The relation $\rightsquigarrow$ defines a directed graph, which we might call the "influence network". Now let $\mathbf{D}^n = [d_{i,j}^{(n)}]$; Thus, $d_{i,j}^{(n)} > 0$ if and only if there is at least one directed path of length n from j to i in the influence network; in this case, $d_{i,j}^{(n)}$ measures the total *indirect* influence which j has on i via such chains of intermediaries. Finally, if $\sum_{n=1}^{\infty} q_n \mathbf{D}^n = [e_{i,j}]_{i,j \in \mathcal{I}}$, then $e_{i,j}$ measures the total influence which j has on i over all possible chains of all possible lengths (weighted by the vector $\mathbf{q}$).

An interesting special case is when $\rightsquigarrow$ is an *acyclic digraph* on $\mathcal{I}$ (that is: a binary relation which is irreflexive, antisymmetric, and whose transitive closure contains no cycles). In this case, the society has a hierarchical structure: there are "opinion leaders" (who are further upstream with respect to $\rightsquigarrow$) and "followers" (who are downstream from the opinion leaders). Informally, "opinion leaders" correspond to pundits, politicians, public intellectuals, and religious authorities, who can influence a large audience of "followers". The deliberative institution will be local as long as the opinion leaders do not have too strong an influence on their followers.

Conclusion

We have shown that a large class of voting rules will converge to the correct solution in a large enough population, even if there is considerable correlation between voters. This suggests, for example, that a large committee of experts can often provide accurate answers to technical questions in science, medicine, or engineering. It also seems to suggest that, under some conditions, modern mass democracies could exhibit a high level of collective epistemic competence. However, before drawing such a conclusion, it is important to recognize that some of our modelling assumptions may be overly optimistic. For example, perhaps the hypotheses of *Identification* and *Asymptotic Determinacy* impute an unrealistically high level of epistemic competence to the average voter. There is now abundant empirical evidence that human beings are subject to systematic cognitive biases, particularly in tasks which involve logical or probabilistic reasoning (Kahneman, 2011). They also overestimate small but spectacular risks (e.g. terrorism), while neglecting threats which are less visible but far more pervasive and hazardous (e.g. antibiotic resistant bacteria). They gravitate towards simple solutions, based on simplistic moral narratives. A more sophisticated theory of epistemic democracy should account for such cognitive biases.

Ironically, the purported epistemic competency of large groups may be self-refuting. By combining the strategic analysis of Austen-Smith and Banks (1996) with the "rational ignorance" of Downs (1957), a voter might decide that there is no reason for her to become informed at all, because the group is going to get the right answer anyways. If enough voters behave this way, then the epistemic competency of the group may be undermined.[17] To counteract such "epistemic free-riding", perhaps we must offer each voter an individual incentive to get the right answer. It is notable that Galton's (1907) original inspiration was a betting pool, not a referendum.

We might also question our assumption that the set $\mathcal{S}$ of social alternatives can be identified one-for-one with the possible states of the world. In reality, the alternatives in $\mathcal{S}$ are generated by some murky and epistemically dubious political process, and it is possible that *none* of these alternatives correctly describes the actual state of the world. Suppose $\mathcal{S} = \{f, t_1, t_2, t_3, t_4\}$, where f is a completely false theory, while the theories t_1, t_2, t_3, t_4 are each somewhat flawed but "mostly true". Then even in a society of highly competent voters, where 75% select one of the "mostly true" theories, the false theory f might win

[17] But see Martinelli (2006) for a counterargument.

a plurality vote through vote-splitting, contradicting the predictions of Example 4.2(b). And this assumes that $\mathcal{S}$ consists of clear descriptions of possible worlds at all; in some cases, the statements in $\mathcal{S}$ may be ambiguous or even meaningless.

Furthermore, in many collective decisions, epistemic questions are inextricable intertwined with questions of ethical values and/or individual preferences. In real life, elections and referenda rarely boil down to objective, "purely factual" questions of the kind considered in this paper. Even when it is possible to isolate such "purely factual" questions in policy debates, many voters cling to the position which they find the most ideologically congenial, rather than the position which is best supported by the available scientific evidence.

It is also possible that modern mass democracies actually exhibit a much higher degree of voter correlation than we allowed in our models. The hypothesis of *Asymptotically Weak Average Correlation* is consistent with a world where most correlations arise from "local" interactions —e.g. through links in a social network, or via person-to-person deliberation. It is even consistent with an Internet-saturated world, where voters are influenced by bloggers and other social media celebrities whose audiences follow a power law distribution (Example 6.1). However, these models assume that the process which generates the social network topology is entirely independent of the process which generates the voters' opinions. In practice, these two processes are highly interdependent, because people preferentially affiliate with other people who share their opinions. This can lead to the formation of "echo chambers", within which deliberation actually *reduces* epistemic competency, by reinforcing voters' ideological biases and cultivating manichean extremism (Sunstein, 2003, 2009). A properly functioning epistemic democracy needs mechanisms to prevent the formation of such echo chambers. Thus, many proposals for deliberative democracy emphasize randomly selected juries or deliberative assemblies (Leib, 2004; Fishkin, 2009).

Finally, the growing concentration of media ownership in modern societies means that most voters get most of their information about the world from a very small number of genuinely independent sources. If we take the epistemic view of democracy seriously, then one possible policy implication is that governments should be much more aggressive in preventing the burgeoning oligopolization of radio, television and print media.

Acknowledgements. This research was supported by Labex MME-DII (ANR11-LBX-0023-01). It was also partly supported by COST Action IC1205 on Computational Social Choice. I am grateful to the participants of the October 2014 COST IC1205 conference at Lucian Blaga University in Sibiu, Romania for their valuable feedback, especially Ulle Endriss and Jérôme Lang. I also thank Franz Dietrich and Stéphane Zuber for their comments. Finally, I would like to thank an anonymous referee for an exceptionally careful, thoughtful, and helpful report, which caught several mistakes and greatly improved the paper. None of these people are responsible for any remaining errors.

A Appendix

The following result will be used in our analysis of Example 4.5 below. The proof is well-known, but it is short, so we include it for completeness.

Lemma A.1 *Fix* $\mathbf{p} \in \Delta(\mathcal{S})$. *Define* $F_{\mathbf{p}} : \Delta(\mathcal{S}) \longrightarrow \mathbb{R}$ *by* $F_{\mathbf{p}}(\mathbf{q}) := \sum_{s \in \mathcal{S}} p_s \log[q_s]$.[18] *Then* $\operatorname{argmax}(F) = \mathbf{p}$.

Proof: We use the method of Lagrange multipliers. Let $\mathbf{1} \in \mathbb{R}^{\mathcal{S}}$ be the constant 1 vector Note that $\Delta(\mathcal{S}) := \{\mathbf{r} \in \mathbb{R}^{\mathcal{S}}_{+};\ \ \mathbf{1} \bullet \mathbf{r} = 1\}$. Thus, if an interior maximum $\mathbf{q}^*$ exists, it must satisfy the first-order condition that $\nabla F_{\mathbf{p}}(\mathbf{q}^*) = c\,\mathbf{1}$ for some constant $c \in \mathbb{R}$.

Now, for all $s \in \mathcal{S}$, we have $\partial_s F(\mathbf{q}) = p_s/q_s$. Thus, $\nabla F_{\mathbf{p}}(\mathbf{q}) = c\,\mathbf{1}$ if and only if $p_s = c\,q_s$ for all $s \in \mathcal{S}$. Since $\mathbf{p}$ and $\mathbf{q}$ are both probability vectors, this can happen only if $c = 1$ and $\mathbf{p} = \mathbf{q}$. Thus, the unique critical point of $F_{\mathbf{p}}$ is at $\mathbf{p}$ itself.

Finally, observe that $F_{\mathbf{p}}$ is concave (indeed, $\partial_t \partial_s F_{\mathbf{p}} = 0$ if $s \neq t$, whereas $\partial_s^2 F_{\mathbf{p}}(\mathbf{q}) = -p_s/q_s^2 < 0$, so the Hessian is a negative diagonal matrix, hence negative-definite everywhere). Thus, this critical point is a maximum. □

The next result deals with the unproved assertions in Example 4.5.

Proposition A.2 *Let* $\mathcal{S}$ *be finite, let* $p : \mathcal{S} \longrightarrow \Delta(\mathcal{S})$ *be any error model, let* $\eta > 0$, *and define* $F^p_{\log}$ *and* $\mathcal{P}_{p,\eta}$ *as in Example 4.5.*

(a) *If* $p(t|s) > 0$ *for all* $t, s \in \mathcal{S}$, *and* $\eta < \min\{p(t|s);\ \ s, t \in \mathcal{S}\}$, *then* $\mathcal{P}_{p,\eta}$ *satisfies* Minimal Determinacy.

(b) *Let* $\epsilon > 0$, *and define* $\mathcal{C}'_\epsilon$ *as in Example 4.5. If* ϵ *and* η *are small enough, then* $\mathcal{P}_{p,\eta}$ *satisfies* Identification *with respect to* $F^p_{\log}$ *and* $\mathcal{C}'_\epsilon$.

Proof: (a) Let $M := \min\{p(t|s);\ t, s \in \mathcal{S}\}$; then $M > 0$, because $\mathcal{S}$ is finite, and $p(t|s) > 0$ for all $t, s \in \mathcal{S}$. Let $L := |\log(M)|$. Then $L < \infty$, and we have $|v^t_s| \leq L$ for all $s, t \in \mathcal{S}$. Thus, $\|\mathbf{v}^t\|^2 \leq L^2\,|S|$ for all $t \in \mathcal{S}$. Thus, $\operatorname{var}[\rho(t)] \leq L^2\,|S|$ for all $t \in \mathcal{S}$. Thus, *Minimal Determinacy* is satisfied.

(b) For all $s \in \mathcal{S}$, recall that $\mathcal{C}^\epsilon_s := \{\mathbf{c} \in \mathcal{C};\ \ c_s \geq c_t + \epsilon$ for all $t \neq s\}$. Suppose p is the error model of a voter. Then for any $s, t \in \mathcal{S}$, we have $\rho(\mathbf{v}^t|s) = p(t|s)$. Thus, $\mathbb{E}[\rho(s)] = \sum_{t \in \mathcal{S}} p(t|s)\mathbf{v}^t = (w_r(s))_{r \in \mathcal{S}}$, where, for all $r \in \mathcal{S}$, $w_r(s) = \sum_{t \in \mathcal{S}} p(t|s) v^t_r = \sum_{t \in \mathcal{S}} p(t|s)\ \log[p(t|r)]$.

We will first construct some $\epsilon_0 > 0$ such that $\mathbb{E}[\rho(s)] \in \mathcal{C}^{\epsilon_0}_s$. To do this, we must show that

$$\sum_{t \in \mathcal{S}} p(t|s)\ \log[p(t|s)] \quad \geq \quad \epsilon_0 + \sum_{t \in \mathcal{S}} p(t|s)\ \log[p(t|r)], \quad \text{for all } r \neq s. \tag{A1}$$

[18]The function $-F_{\mathbf{p}}(\mathbf{q})$ is sometimes called the *cross-entropy* of $\mathbf{p}$ and $\mathbf{q}$.

Now, for any $s \in \mathcal{S}$, define $\mathbf{p}^s$ by setting $p^s_t := p(t|s)$ for all $t \in \mathcal{S}$. Then Lemma A.1 says that $F_{\mathbf{p}^s}(\mathbf{p}^s) > F_{\mathbf{p}^s}(\mathbf{q})$ for all $\mathbf{q} \in \Delta(\mathcal{S}) \setminus \{\mathbf{p}^s\}$. In particular, this implies that $F_{\mathbf{p}^s}(\mathbf{p}^s) > F_{\mathbf{p}^s}(\mathbf{p}^t)$ for all $t \neq s$. Let $\epsilon_0 := F_{\mathbf{p}^s}(\mathbf{p}^s) - \max\{F_{\mathbf{p}^s}(\mathbf{p}^t);\ t \in \mathcal{S} \setminus \{s\}\}$. Then $\epsilon_0 > 0$, because $\mathcal{S}$ is finite. We have $F_{\mathbf{p}^s}(\mathbf{p}^s) \geq F_{\mathbf{p}^s}(\mathbf{p}^t) + \epsilon_0$ for all $t \neq s$. This implies condition (A1).

Now, let $0 < \epsilon < \epsilon_0$. If $\eta > 0$ is small enough, then by continuity we will have

$$\sum_{t\in\mathcal{S}} p'(t|s)\,\log[p'(t|s)] \quad \geq \quad \epsilon + \sum_{t\in\mathcal{S}} p'(t|s)\,\log[p'(t|r)], \quad \text{for all } r \neq s. \tag{A2}$$

for all error models p' such that $|p'(t|s) - p(t|s)| < \eta$ for all $s, t \in \mathcal{S}$. Thus implies that $\mathcal{P}_{p,\eta}$ satisfies *Identification* with respect to $F^p_{\log}$ and $\mathcal{C}'_\epsilon$. □

Proposition 5.2 is just a special case of Theorem 5.3 when $\mathcal{S}$ is a finite set with the discrete topology. Likewise, Proposition 4.1 is just a special case of Proposition 4.3 when $\mathcal{S}$ is a finite set with the discrete topology. It remains to prove Theorem 5.3.

Proof of Theorem 5.3. Let $F = (\mathbb{V}, \mathcal{V}, f)$ be a mean partition rule. Let $\mathcal{C}$ be the closed convex hull of $\mathcal{V}$, let $\mathcal{C}' \subseteq \mathcal{C}$ and $\delta > 0$ be as in (M3), and let f_0 be the (uniformly continuous) restriction of f to $\mathcal{C}^\delta$. Fix $s \in \mathcal{S}$, and let $\mathcal{U} \subset \mathcal{S}$ be an open set containing s. For all $I \in \mathbb{N}$, let $\rho \in \mathcal{R}_I$, let $\mathbf{V}_I := (\mathbf{v}_i)_{i\in\mathcal{I}}$ be a $\rho(s)$-random profile of votes (where $|\mathcal{I}| = I$), and let $\overline{\mathbf{v}}_I := \dfrac{1}{I}\displaystyle\sum_{i\in\mathcal{I}} \mathbf{v}_i$ be their average. We claim $\displaystyle\lim_{I\to\infty} \text{Prob}[f(\overline{\mathbf{v}}_I) \in \mathcal{U}] = 1$.

Claim 1: *Let* $\widehat{\mathbf{v}}_I := \mathbb{E}[\overline{\mathbf{v}}_I]$. *Then* $\widehat{\mathbf{v}}_I \in f_0^{-1}\{s\}$.

Proof: $\mathbb{E}(\overline{\mathbf{v}}_I) = \mathbb{E}\left(\frac{1}{I}\sum_{i\in\mathcal{I}} \mathbf{v}_i\right) = \frac{1}{I}\sum_{i\in\mathcal{I}} \mathbb{E}(\mathbf{v}_i)$. By *Identification*, we have $\mathbb{E}(\mathbf{v}_i) \in f_0^{-1}\{s\}$ for all i. But $f_0^{-1}\{s\}$ is convex by (M4). The claim follows. ◇ `Claim 1`

Claim 2: $\displaystyle \text{var}(\overline{\mathbf{v}}_I) \leq \frac{1}{I}\,\sigma(I) + \frac{1(I-1)}{I}\,\kappa(I).$

Proof: For any $i \in \mathcal{I}$, let $\widehat{\mathbf{v}}_i := \mathbb{E}[\mathbf{v}_i]$. Then as we saw in the proof of Claim 1, $\widehat{\mathbf{v}}_I = \dfrac{1}{I}\displaystyle\sum_{i\in\mathcal{I}} \widehat{\mathbf{v}}_i$. Thus,

$$\overline{\mathbf{v}}_I - \widehat{\mathbf{v}}_I \quad = \quad \frac{1}{I}\sum_{i\in\mathcal{I}} \mathbf{v}_i - \frac{1}{I}\sum_{i\in\mathcal{I}} \widehat{\mathbf{v}}_i \quad = \quad \frac{1}{I}\sum_{i\in\mathcal{I}} (\mathbf{v}_i - \widehat{\mathbf{v}}_i) \quad = \quad \frac{1}{I}\sum_{i\in\mathcal{I}} \mathbf{e}_i,$$

where, for all $i \in \mathcal{I}$, we define $\mathbf{e}_i := \mathbf{v}_i - \widehat{\mathbf{v}}_i$. Thus,

$$\|\overline{\mathbf{v}}_I - \widehat{\mathbf{v}}_I\|^2 \quad = \quad \left\langle \frac{1}{I}\sum_{i\in\mathcal{I}} \mathbf{e}_i, \frac{1}{I}\sum_{j\in\mathcal{I}} \mathbf{e}_j \right\rangle \quad = \quad \frac{1}{I^2}\sum_{i,j\in\mathcal{I}} \langle \mathbf{e}_i, \mathbf{e}_j \rangle,$$

Thus, if $\mathbf{B}$ is the covariance matrix of $\rho(s)$, then

$$\begin{aligned}
\mathrm{var}(\overline{\mathbf{v}}_I) &= \mathbb{E}\left[\|\overline{\mathbf{v}}_I - \widehat{\mathbf{v}}_I\|^2\right] \quad = \quad \frac{1}{I^2}\sum_{i,j\in\mathcal{I}} \mathbb{E}\left[\langle \mathbf{e}_i, \mathbf{e}_j\rangle\right] \\
&= \frac{1}{I^2}\sum_{i,j\in\mathcal{I}} \mathrm{cov}(\mathbf{v}_i, \mathbf{v}_j) \quad = \quad \frac{1}{I^2}\sum_{i\in\mathcal{I}} \mathrm{var}(\mathbf{v}_i) + \frac{1}{I^2}\sum_{\substack{i,j\in\mathcal{I}\\ i\neq j}} \mathrm{cov}(\mathbf{v}_i, \mathbf{v}_j) \\
&\underset{(*)}{=} \frac{1}{I}\sigma(\mathbf{B}) + \frac{1(I-1)}{I}\kappa(\mathbf{B}) \quad \underset{(\dagger)}{\leq} \quad \frac{1}{I}\sigma(I) + \frac{1(I-1)}{I}\kappa(I),
\end{aligned}$$

as claimed. Here, $(*)$ is by the defining equations (2), and $(\dagger)$ is by definition of $\sigma(I)$ and $\kappa(I)$. $\diamond$ `Claim 2`

Now, f_0 is uniformly continuous on $\mathcal{C}^\delta$, by (M3). Thus, there exists $\eta > 0$ with the following property:

$$\text{For all } \widehat{\mathbf{c}} \in f_0^{-1}\{s\} \text{ and all } \mathbf{c} \in \mathcal{C}^\delta, \text{ if } \|\mathbf{c} - \widehat{\mathbf{c}}\| < \eta, \text{ then } \ f_0(\mathbf{c}) \in \mathcal{U}. \tag{A3}$$

Meanwhile, Claim 1 says that $\widehat{\mathbf{v}}_I \in f_0^{-1}\{s\}$; thus, $\widehat{\mathbf{v}}_I \in \mathcal{C}'$. Thus, if $\mathbf{c} \in \mathcal{C}$, and $\|\mathbf{c} - \widehat{\mathbf{v}}_I\| < \delta$, then $\mathbf{c} \in \mathcal{C}^\delta$. Let $\epsilon := \min\{\delta, \eta\}$; then $\epsilon > 0$, and for any $\mathbf{c} \in \mathcal{C}$, if $\|\mathbf{c} - \widehat{\mathbf{v}}_I\| < \epsilon$, then property (A3) implies that $f(\mathbf{c}) \in \mathcal{U}$. In particular, this holds if $\mathbf{c} = \overline{\mathbf{v}}_I$. Thus,

$$\begin{aligned}
\mathrm{Prob}\left[f(\overline{\mathbf{v}}_I) \notin \mathcal{U}\right] \quad &\leq \quad \mathrm{Prob}\left[\|\widehat{\mathbf{v}}_I - \overline{\mathbf{v}}_I\| > \epsilon\right] \quad \underset{(*)}{\leq} \quad \frac{\mathrm{var}(\overline{\mathbf{v}}_I)}{\epsilon^2} \\
&\underset{(\dagger)}{\leq} \quad \frac{1}{\epsilon^2}\left(\frac{1}{I}\,\sigma(I) + \frac{I-1}{I}\,\kappa(I)\right) \quad \xrightarrow[I\to\infty]{(\diamond)} \quad 0,
\end{aligned}$$

as desired. Here $(*)$ is by the normed vector space version of Chebyshev's inequality, and $(\dagger)$ is by Claim 2. Finally, $(\diamond)$ is because *Asymptotically Minimal Determinacy* says that $\frac{1}{I}\,\sigma(I)\xrightarrow[I\to\infty]{}0$, while *Asymptotically Weak Average Covariance* says that $\kappa(I)\xrightarrow[I\to\infty]{}0$. $\square$

Remark A.3. Fix $s \in \mathcal{S}$, and write ρ_s for $\rho(s)$. We have defined $\mathbb{E}[\rho(s)]$ to be the *expected value* of a ρ_s-random variable. Formally, this is the following $\mathbb{V}$-valued integral:

$$\mathbb{E}[\rho(s)] \quad := \quad \int_{\mathcal{V}} \mathbf{v}\ \mathrm{d}\rho_s[\mathbf{v}]. \tag{A4}$$

If $\mathbb{V}$ is a finite-dimensional vector space, then the integral in (A4) is defined in the standard way, by simply computing the Lebesgue integral of each coordinate. More generally, when $\mathbb{V}$ is possibly infinite-dimensional, we interpret (A4) as the *Bochner integral* of the identity function $I : \mathbb{V}\longrightarrow\mathbb{V}$ with respect to the measure ρ_s. (This agrees with the coordinatewise Lebesgue integral when $\mathbb{V}$ is finite-dimensional.) To be precise, suppose ρ_s is defined on a sigma-algebra $\mathfrak{B}$ of subsets of $\mathbb{V}$. A $\mathfrak{B}$-measurable function $f : \mathbb{V}\longrightarrow\mathbb{V}$ is $\mathfrak{B}$-*simple* if it takes only finitely many values. If f takes the values $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_N$ on the $\mathfrak{B}$-measurable sets $\mathcal{B}_1, \ldots, \mathcal{B}_N$ respectively, then we define

$$\int_{\mathbb{V}} f \; \mathrm{d}\rho_s \quad := \quad \sum_{n=1}^{N} \rho_s[\mathcal{B}_n] \, \mathbf{v}_n. \tag{A5}$$

Now let $\{f_n\}_{n=1}^{\infty}$ be a sequence of $\mathfrak{B}$-simple functions from $\mathbb{V}$ to itself, such that

$$\lim_{n\to\infty} f_n(\mathbf{v}) \;=\; \mathbf{v} \quad \text{for } \rho_s\text{-almost all } \mathbf{v} \in \mathbb{V}, \tag{A6}$$

and

$$\lim_{n\to\infty} \int_{\mathbb{V}} \|I - f_n\| \; \mathrm{d}\rho_s \quad = \quad 0. \tag{A7}$$

Then the *Bochner integral* is defined as the limit

$$\int_{\mathcal{V}} \mathbf{v} \; \mathrm{d}\rho_s[\mathbf{v}] \quad := \quad \lim_{n\to\infty} \int_{\mathbb{V}} f_n \; \mathrm{d}\rho_s, \tag{A8}$$

where each of the integrals on the right hand side of (A8) is defined as in (A5). If the limit (A8) exists, then it is independent of the particular sequence of simple functions we use to approximate the identity function (Aliprantis and Border, 2006, Lemma 11.41, p.425).

Thus, for the expected value (A4) to be well-defined as a Bochner integral, two conditions must be satisfied. First, we need a sequence $\{f_n\}_{n=1}^{\infty}$ of $\mathfrak{B}$-simple functions satisfying convergence conditions (A6) and (A7). Second, we need the limit (A8) to exist.

There is a sequence $\{f_n\}_{n=1}^{\infty}$ satisfying (A6) if and only there is a separable closed subspace $\mathbb{V}_0$ of $\mathbb{V}$ such that $\rho[\mathbb{V}_0] = 1$ (Aliprantis and Border, 2006, Lemma 11.37). (Clearly, this holds if $\mathbb{V}$ itself is separable.) Suppose $\mathfrak{B}$ is the Borel sigma-algebra induced by the norm topology on $\mathbb{V}$. Then we can obtain a sequence that also satisfies (A7) if, for any $\epsilon > 0$, there is some compact subset $\mathcal{K} \subset \mathbb{V}_0$ such that $\int_{\mathcal{K}^{\complement}} \|\mathbf{v}\| \; \mathrm{d}\rho_s[\mathbf{v}] < \epsilon$. (This is easy to verify.) In particular, this holds if ρ_s is "almost-compactly supported", meaning that there is a norm-bounded subset $\mathcal{B} \subset \mathbb{V}_0$ with $\rho_s[\mathcal{B}] = 1$, and for any $\epsilon > 0$, there is some compact subset $\mathcal{K} \subseteq \mathcal{B}$ such that $\rho_s[\mathcal{K}] > 1 - \epsilon$. Finally, the limit (A8) is guaranteed to exist if $\mathbb{V}$ is a Hilbert space —i.e. the inner product metric is Cauchy-complete (Aliprantis and Border, 2006, Lemma 11.41). But the limit (A8) could exist even when $\mathbb{V}$ is not a Hilbert space. For example, suppose ρ_s is a finite sum of point masses. Let f be a simple function such that $f(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v}$ where ρ_s has a point mass. If we define $f_n := f$ for all $n \in \mathbb{N}$, then the sequence $\{f_n\}_{n=1}^{\infty}$ trivially satisfies (A6) and (A7), and the limit (A8) trivially exists. A Hilbert space structure is not required for any of our main results. So we have not assumed that $\mathbb{V}$ is a Hilbert space in condition (M1).

Proof of Proposition 5.4. Let $\mathbb{V}$ be a vector space with inner product $\langle \_, \_ \rangle$, and suppose that $\mathfrak{R}$ is an identifiable culture on $\mathbb{V}$. For all $s \in \mathcal{S}$ and $j \in \mathcal{I}$, let $\mathcal{V}_{s,j} \subseteq \mathbb{V}$ be a subset such that, if $(\mathbf{v}_i)_{i\in\mathcal{I}}$ is a $\rho(s)$-random profile, then $\mathrm{Prob}[\mathbf{v}_j \in \mathcal{V}_{s,j}] = 1$. Then let

$$\mathcal{V} \quad := \quad \bigcup_{s\in\mathcal{S}} \bigcup_{i\in\mathcal{I}} \mathcal{V}_{i,s}.$$

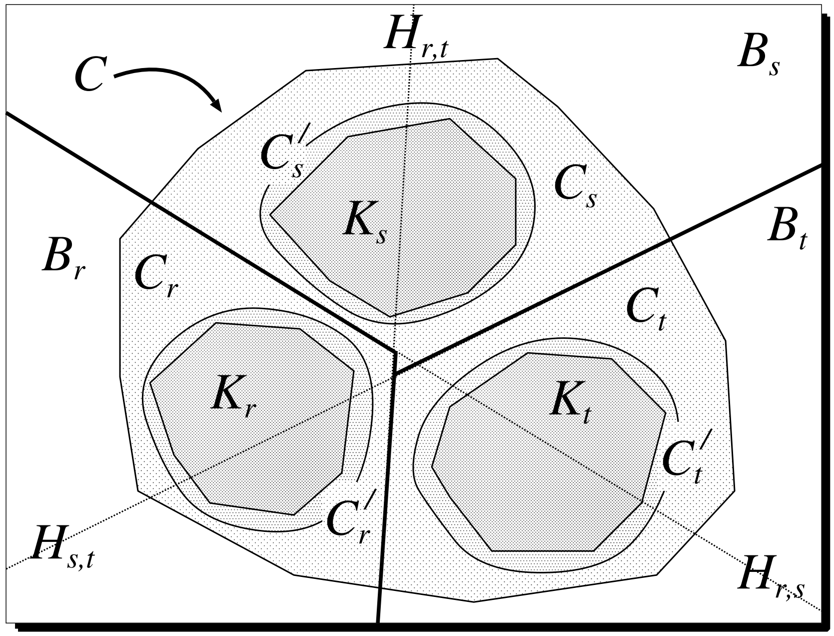

Figure 9: The proof of Proposition 5.4. (In this example, $\mathcal{S} = \{r, s, t\}$ and $\mathbb{V} = \mathbb{R}^2$.)

Let $\mathcal{C}$ be the closed, convex hull of $\mathcal{V}$. Let $\{\mathcal{K}_s\}_{s\in\mathcal{S}}$ be the compact, convex subsets of $\mathbb{V}$ in the definition of identifiability. Without loss of generality, we can suppose that these are subsets of $\mathcal{C}$ (otherwise, replace each $\mathcal{K}_s$ with $\mathcal{K}_s \cap \mathcal{C}$, which is also compact and convex.)

For any distinct $r, s \in \mathcal{S}$, let $\epsilon_{r,s}$ be the minimum distance between $\mathcal{K}_r$ and $\mathcal{K}_s$; then $\epsilon_{r,s} > 0$ because $\mathcal{K}_r$ and $\mathcal{K}_s$ are compact and disjoint. Let

$$\epsilon \quad := \quad \frac{1}{4} \min_{\substack{r,s\in\mathcal{S} \\ r\neq s}} \epsilon_{r,s}.$$

Then $\epsilon > 0$ because $\mathcal{S}$ is finite. For any distinct $r, s \in \mathcal{S}$, the convex sets $\mathcal{K}_s$ and $\mathcal{K}_r$ are disjoint, so there is a hyperplane $\mathcal{H}_{r,s}$ which passes between them. Furthermore, we can arrange for this hyperplane to have distance $\epsilon_{r,s}/2$ from each of $\mathcal{K}_r$ and $\mathcal{K}_s$, as shown in Figure 9. For any $s \in \mathcal{S}$, let $\mathcal{A}_s \subseteq \mathbb{V}$ be the closed, convex set supported by all the hyperplanes $\{\mathcal{H}_{r,s};\ \ r \in \mathcal{S} \setminus \{s\}\}$. (Thus, $\mathcal{A}_r \cap \mathcal{A}_s \subseteq \mathcal{H}_{r,s}$, for any distinct $r, s \in \mathcal{S}$.) These convex sets may not cover all of $\mathbb{V}$. (For example, the small central triangle in Figure 9 is not covered by any $\mathcal{A}_s$.) So we construct the sets $\mathcal{B}_s$ (for all $s \in \mathcal{S}$) by attaching any uncovered part of $\mathbb{V}$ to $\mathcal{A}_s$, for some arbitrary $s \in \mathcal{S}$, so that $\mathbb{V} = \bigcup_{s\in\mathcal{S}} \mathcal{B}_s$. Let $\mathcal{C}'_s := \{\mathbf{c} \in \mathcal{C};\ \ d(\mathbf{v}, \mathcal{K}_s) < \epsilon\}$, as shown in Figure 9; then $\mathcal{C}'_s$ is contained in the interior of $\mathcal{A}_s$, and thus, in the interior of $\mathcal{B}_s$.

Suppose $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$. For all $n \in [1 \ldots N]$, define $\mathcal{B}^*_{s_n} := \mathcal{B}_{s_n} \setminus (\mathcal{B}_{s_1} \cup \cdots \cup \mathcal{B}_{s_{n-1}})$. Thus, the sets $\mathcal{B}^*_{s_1}, \ldots, \mathcal{B}^*_{s_N}$ form a partition of $\mathbb{V}$. Thus, if we define $\mathcal{C}_s := \mathcal{B}^*_s \cap \mathcal{C}$ for all $s \in \mathcal{S}$, then the sets $\mathcal{C}_{s_1}, \ldots, \mathcal{C}_{s_N}$ form a partition of $\mathcal{C}$. For each $s \in \mathcal{S}$, $\mathcal{C}'_s$ is in the interior of $\mathcal{B}^*_s$, by construction; thus, $\mathcal{C}'_s$ is in the interior of $\mathcal{C}_s$. Indeed, every point of

$\mathcal{C}'_s$ is at least ϵ-distant from $\mathcal{H}_{r,s}$ (for all $r \in \mathcal{S}$), and thus, is at least ϵ-distant from the boundary of $\mathcal{C}_s$. Now define the function $f : \mathcal{C} \longrightarrow \mathcal{S}$ by setting $f(\mathbf{c}) := s$ for all $\mathbf{c} \in \mathcal{C}_s$, for all $s \in \mathcal{S}$. Then let $F := (\mathbb{V}, \mathcal{V}, f)$. To see that F is a mean partition rule, let $\mathcal{C}' := \bigcup_{s\in\mathcal{S}} \mathcal{C}'_s$ and let $\delta := \epsilon/2$. Then $\mathcal{C}'$ and δ satisfy properties (M3) and (M4) (because $\mathcal{C}' \cap \mathcal{C}_s = \mathcal{C}'_s$ for all $s \in \mathcal{S}$). Finally, by construction, F satisfies the axiom *Identification* with respect to $\mathfrak{R}$. □

Proof of Proposition 6.5. Let $\beta : \mathbb{N} \longrightarrow \mathbb{R}^+$ and $\gamma : \mathbb{N} \longrightarrow [0, \infty]$ be functions satisfying the inequality (9), such that $\mathfrak{B}$ exhibits β- covariance decay relative to $\mathfrak{N}$, and $\mathfrak{N}$ has sublinear average γ-degree growth. Let $I \in \mathbb{N}$, let $\mathbf{B} \in \mathcal{B}_I$, and let $(\mathcal{I}, \sim)$ be a graph in $\mathcal{N}_I$ such that $\mathbf{B}_I$ exhibits β-decay for $(\mathcal{I}, \sim)$. Let $M := \beta(0)$; then *Minimal Determinacy* is automatically satisfied, because $|b_{i,i}| \leq \beta(0)$ for all $i \in \mathcal{I}$. It remains to prove *Asymptotically weak average covariance* . Let $C := \sum_{n=1}^{\infty} \gamma(n)\, \beta(n)$; then C is finite by inequality (9). We have:

$$
\begin{aligned}
\kappa(\mathbf{B}) \quad &= \quad \frac{1}{I(I-1)} \sum_{\substack{i,j\in\mathcal{I} \\ i\neq j}}^{I} b_{i,j} \quad = \quad \frac{1}{I(I-1)} \sum_{i\in\mathcal{I}} \sum_{r=1}^{\infty} \sum_{\substack{j\in\mathcal{I} \\ d(i,j)=r}} b_{i,j} \\
&\underset{\text{(a)}}{\leq} \quad \frac{1}{I(I-1)} \sum_{i\in\mathcal{I}} \sum_{r=1}^{\infty} \sum_{\substack{j\in\mathcal{I} \\ d(i,j)=r}} \beta(r) \quad = \quad \frac{1}{I(I-1)} \sum_{i\in\mathcal{I}} \sum_{r=1}^{\infty} \beta(r)\ \deg^r(i, \sim) \\
&\underset{\text{(b)}}{\leq} \quad \frac{1}{I(I-1)} \sum_{i\in\mathcal{I}} \sum_{r=1}^{\infty} \beta(r)\, \gamma(r)\ \deg^\gamma(i, \sim) \\
&= \quad \frac{1}{I(I-1)} \sum_{i\in\mathcal{I}} \deg^\gamma(i, \sim) \left(\sum_{r=1}^{\infty} \beta(r)\, \gamma(r) \right) \quad \underset{\text{(c)}}{=} \quad \frac{1}{I(I-1)} \sum_{i\in\mathcal{I}} \deg^\gamma(i, \sim) \cdot C \\
&\underset{\text{(d)}}{=} \quad \frac{C}{(I-1)}\ \mathrm{avedeg}^\gamma(\mathcal{I}, \sim) \quad \underset{\text{(e)}}{\leq} \quad \frac{C}{(I-1)}\ \overline{\mathrm{avedeg}}^\gamma(\mathcal{N}_I).
\end{aligned}
$$

Here, inequality (a) is because $\mathbf{B}$ exhibits β-decay for $(\mathcal{I}, \sim)$, while inequality (b) is by defining formula (4). Equality (c) is by definition of C, and equality (d) is by defining formula (5). Inequality (e) is by defining formula (6).

This inequality holds for all matrices $\mathbf{B} \in \mathcal{B}_I$. It follows that

$$
\kappa(I) \quad \leq \quad \frac{C}{(I-1)}\ \overline{\mathrm{avedeg}}^\gamma(\mathcal{N}_I) \quad \xrightarrow[I\to\infty]{} \quad 0,
$$

as desired, where the last step is by the limit equation (7). □

Proof of Proposition 7.1. Let $\mathfrak{R}' := \mathfrak{D} \odot \mathfrak{R}$. We must verify the three conditions for $\mathfrak{R}'$ to be sagacious. First, we will show that $\mathfrak{R}'$ satisfies *Identification.* Suppose the true state of nature is s. Fix $I \in \mathbb{N}$, and let $\mathcal{I} := [1 \ldots I]$. Let $\mathcal{R}'_I = \mathcal{D}_I \odot \mathcal{R}_I$, let $\rho' \in \mathcal{R}'_I$, and let $\mathbf{V}' = (\mathbf{v}'_i)_{i=1}^I$ be a $\rho'(s)$-random profile. Then there exists a collective behaviour model $\rho \in \mathcal{R}_I$, and an influence matrix $\mathbf{D} \in \mathcal{D}_I$ such that $\rho' = \mathbf{D} \odot \rho$. Suppose $\mathbf{D} = [d_{i,j}]_{i,j \in \mathcal{I}}$. Thus, for all $i \in \mathcal{I}$, we have

$$\mathbf{v}'_i \quad := \quad \sum_{j=1}^{I} d_{i,j}\, \mathbf{v}_j,$$

where $\mathbf{V} = (\mathbf{v}_i)_{i=1}^I$ is a $\rho(s)$-random profile. Now, $\mathcal{R}_I$ is sagacious, so it satisfies *Identification*; thus, for all $k \in \mathcal{I}$, we have $\mathbb{E}[\mathbf{v}_k] \in F^{-1}\{s\} \cap \mathcal{C}'$. Thus, for all $i \in \mathcal{I}$, we have

$$\mathbb{E}[\mathbf{v}'_i] \;=\; \mathbb{E}\left[\sum_{k \in \mathcal{I}} d_{i,k}\, \mathbf{v}_k\right] \;=\; \sum_{k \in \mathcal{I}} d_{i,k}\, \mathbb{E}[\mathbf{v}_k] \in F^{-1}\{s\} \cap \mathcal{C}',$$

because $F^{-1}\{s\} \cap \mathcal{C}'$ is convex by (M4), and $\sum_{k \in \mathcal{I}} d_{i,k} = 1$ (because $\mathbf{D}$ is a stochastic matrix). Thus, *Identification* is satisfied.

It remains to show that $\mathfrak{R}'$ satisfies *Asymptotically weak average covariance* and *Asymptotic Determinacy*. Since the culture $\mathfrak{R}$ is sagacious, it already satisfies these properties. For all $I \in \mathbb{N}$, let $\sigma(I)$ and $\kappa(I)$ be as defined in the statements of these conditions. Let $s \in \mathcal{S}$, $\mathbf{D} \in \mathcal{D}_I$, $\rho \in \mathcal{R}_I$, $\rho' = \mathbf{D} \odot \rho$, $\mathbf{V}'$, $\mathbf{V}$, etc. be as defined in the proof of *Identification* above. Let $\mathbf{B}' = [b'_{i,j}]_{i,j=1}^I$ be the covariance matrix of $\rho'(s)$. That is: $b'_{i,j} := \mathrm{cov}(\mathbf{v}'_i, \mathbf{v}'_j)$, for all $i, j \in \mathcal{I}$. Let D be the modulus of $\mathfrak{D}$ (this is finite because $\mathfrak{D}$ is local).

Claim 1: $\displaystyle \frac{1}{I^2} \sum_{i,j \in \mathcal{I}} b'_{i,j} \;\leq\; D^2 \left(\frac{1}{I}\sigma(I) + \frac{I-1}{I}\kappa(I) \right).$

Proof: Let $\mathbf{B} = [b_{i,j}]_{i,j=1}^I$ be the covariance matrix of $\rho(s)$. Then for all $i, j \in \mathcal{I}$, we have

$$\begin{aligned} b'_{i,j} \;&=\; \mathrm{cov}(\mathbf{v}'_i, \mathbf{v}'_j) \;=\; \mathrm{cov}\left(\sum_{k \in \mathcal{I}} d_{i,k}\, \mathbf{v}_k, \sum_{\ell \in \mathcal{I}} d_{j,\ell}\, \mathbf{v}_\ell \right) \\ &=\; \sum_{k \in \mathcal{I}} \sum_{\ell \in \mathcal{I}} d_{i,k}\, d_{j,\ell}\, \mathrm{cov}(\mathbf{v}_k, \mathbf{v}_\ell) \;=\; \sum_{k,\ell \in \mathcal{I}} d_{i,k}\, d_{j,\ell}\, b_{k,\ell}, \end{aligned} \tag{A9}$$

where the last step is because $\operatorname{cov}(\mathbf{v}_k, \mathbf{v}_\ell) = b_{k,\ell}$. For all $k \in \mathcal{I}$, let $\overline{d}_k := \sum_{i \in \mathcal{I}} d_{i,k}$. Then

$$
\begin{aligned}
\frac{1}{I^2} \sum_{i,j \in \mathcal{I}} b'_{i,j} &\underset{\text{(a)}}{=} \frac{1}{I^2} \sum_{i,j,k,\ell \in \mathcal{I}} d_{i,k}\, d_{j,\ell}\, b_{k,\ell} \;=\; \frac{1}{I^2} \sum_{k,\ell \in \mathcal{I}} \left(\sum_{i \in \mathcal{I}} d_{i,k} \right) \left(\sum_{j \in \mathcal{I}} d_{j,\ell} \right) b_{k,\ell} \\
&= \frac{1}{I^2} \sum_{k,\ell \in \mathcal{I}} \overline{d}_k\, \overline{d}_\ell\, b_{k,\ell} \;\underset{\text{(b)}}{\leq}\; \frac{1}{I^2} \sum_{k,\ell \in \mathcal{I}} D^2\, b_{k,\ell} \\
&= D^2 \left(\frac{1}{I^2} \sum_{k \in \mathcal{I}} b_{k,\ell} + \frac{1}{I^2} \sum_{\substack{k,\ell \in \mathcal{I} \\ k \neq \ell}} b_{k,\ell} \right) \;=\; D^2 \left(\frac{1}{I} \sigma(\mathbf{B}) + \frac{I-1}{I} \kappa(\mathbf{B}) \right) \\
&\underset{\text{(c)}}{\leq} D^2 \left(\frac{1}{I} \sigma(I) + \frac{I-1}{I} \kappa(I) \right).
\end{aligned}
$$

as claimed. Here, (a) is by equation (A9), while (b) is by definition of "modulus". (c) is by the definitions of $\sigma(I)$ and $\kappa(I)$. $\diamondsuit$ `Claim 1`

Now, let $\mathfrak{B}' = (\mathcal{B}'_I)_{I=1}^{\infty}$ be the covariance structure for the culture $\mathfrak{R}'$. Then for any $I \in \mathbb{N}$ and $\mathbf{B}' \in \mathcal{B}'_I$, we can find some $\rho' \in \mathcal{R}'_I$ and $s \in \mathcal{S}$ such that $\mathbf{B}' = \operatorname{cov}[\rho'(s)]$, and thus, Claim 1 applies to $\mathbf{B}'$. However,

$$
\frac{1}{I^2} \sum_{i,j \in \mathcal{I}} b'_{i,j} \;=\; \frac{I-1}{I} \kappa(\mathbf{B}') + \frac{1}{I} \sigma(\mathbf{B}').
$$

Thus, Claim 1 implies that

$$
\frac{I-1}{I} \kappa(I) + \frac{1}{I} \sigma(I) \quad \leq \quad D^2 \left(\frac{1}{I} \sigma(I) + \frac{I-1}{I} \kappa(I) \right) \quad \xrightarrow[I \to \infty]{} \quad 0,
$$

where the last step because $\mathfrak{R}$ satisfies *Asymptotically weak average covariance* and *Asymptotic Determinacy*. Thus, the culture $\mathfrak{R}'$ is sagacious. $\square$

Proof of Proposition 7.2. Let $\mathcal{I} := [1 \ldots I]$. For any $I \times I$ matrix $\mathbf{D}$, let $\|\mathbf{D}\| := \max_{j \in \mathcal{I}} \left(\sum_{i \in \mathcal{I}} d_{i,j} \right)$. Thus, a deliberative institution $\mathfrak{C} = (\mathcal{C}_I)_{I=1}^{\infty}$ is local if there is some constant $C > 0$ such that $\|\mathbf{C}\| \leq C$ for all $\mathbf{C} \in \mathcal{C}_I$ and all $I \in \mathbb{N}$. In particular, if $\mathfrak{D}$ and $\mathfrak{E}$ are local, then there are constants D and E such that $\|\mathbf{D}\| \leq D$ and and $\|\mathbf{E}\| \leq E$ for all $\mathbf{D} \in \mathcal{D}_I$, all $\mathbf{E} \in \mathcal{E}_I$, and all $I \in \mathbb{N}$.

Claim 1: *For any $I \times I$ matrices $\mathbf{D}$ and $\mathbf{E}$, we have $\|\mathbf{D} \cdot \mathbf{E}\| \leq \|\mathbf{D}\| \cdot \|\mathbf{E}\|$.*

Proof: Let $\mathbf{C} = \mathbf{D} \cdot \mathbf{E}$. Thus, for all $i, k \in \mathcal{I}$, $c_{i,k} = \sum_{j \in \mathcal{I}} d_{i,j}\, e_{j,k}$. Thus, for all $k \in \mathcal{I}$, we have

$$\begin{aligned} \sum_{i \in \mathcal{I}} c_{i,k} &= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} d_{i,j}\, e_{j,k} &=& \sum_{j \in \mathcal{I}} \left(\sum_{i \in \mathcal{I}} d_{i,j} \right) e_{j,k} \\ &\leq \sum_{j \in \mathcal{I}} \|\mathbf{D}\| e_{j,k} &=& \|\mathbf{D}\| \sum_{j \in \mathcal{I}} e_{j,k} \quad \leq \quad \|\mathbf{D}\| \cdot \|\mathbf{E}\|. \end{aligned}$$

Thus, $\|\mathbf{D} \cdot \mathbf{E}\| \leq \|\mathbf{D}\| \cdot \|\mathbf{E}\|$, as claimed. $\diamond$ `Claim 1`

Let $\mathcal{C}_I := \{\mathbf{D} \cdot \mathbf{E};\ \ \mathbf{D} \in \mathcal{D}_I \text{ and } \mathbf{E} \in \mathcal{E}_I\}$. It is well-known that the product of two stochastic matrices is a stochastic matrix. (The proof is very similar to Claim 1.) Thus, every element of $\mathcal{C}_I$ is a stochastic matrix. Meanwhile, it follows from Claim 1 that $\|\mathbf{C}\| \leq D\, E$ for all $\mathbf{C} \in \mathcal{C}_I$ and all $I \in \mathbb{N}$. Thus, $\mathfrak{D} \cdot \mathfrak{E}$ is also local.

Now let $q, q' \in [0, 1]$ such that $q + q' = 1$.

Claim 2: *For any $I \times I$ matrices $\mathbf{D}$ and $\mathbf{E}$, we have $\|q\mathbf{D} + q'\mathbf{E}\| \leq q\|\mathbf{D}\| + q'\|\mathbf{E}\|$.*

Proof: Let $\mathbf{C} = q\mathbf{D} + q'\mathbf{E}$. Thus, for all $i, j \in \mathcal{I}$, $c_{i,j} = q\, d_{i,j} + q'\, e_{i,j}$. Thus, for all $j \in \mathcal{I}$, we have

$$\sum_{i \in \mathcal{I}} c_{i,j} \;=\; \sum_{i \in \mathcal{I}} (q\, d_{i,j} + q'\, e_{i,j}) \;=\; q \sum_{i \in \mathcal{I}} d_{i,j} + q' \sum_{i \in \mathcal{I}} e_{i,j} \;\leq\; q\, \|\mathbf{D}\| + q'\|\mathbf{E}\|.$$

Thus, $\|q\mathbf{D} + q'\mathbf{E}\| \leq q\|\mathbf{D}\| + q'\|\mathbf{E}\|$, as claimed. $\diamond$ `Claim 2`

Let $\mathcal{C}_I := \{q\,\mathbf{D} + q'\,\mathbf{E};\ \mathbf{D} \in \mathcal{D}_I \text{ and } \mathbf{E} \in \mathcal{E}_I\}$. It is well-known that the convex combination of two stochastic matrices is a stochastic matrix. (The proof is very similar to Claim 2.) Thus, every element of $\mathcal{C}_I$ is a stochastic matrix. Meanwhile, it follows from Claim 2 that $\|\mathbf{C}\| \leq q\, D + q'\, E$ for all $\mathbf{C} \in \mathcal{C}_I$ and all $I \in \mathbb{N}$. Thus, $q\mathfrak{D} + q'\mathfrak{E}$ is also local. $\square$

References

Albert, R., Jeong, H., Barabási, A.-L., 1999. Diameter of the world-wide web. Nature 401 (6749), 130–131.

Aliprantis, C. D., Border, K. C., 2006. Infinite dimensional analysis: A hitchhiker's guide, 3rd Edition. Springer, Berlin.

Austen-Smith, D., Banks, J., 1996. Information aggregation, rationality, and the Condorcet jury theorem. American Political Science Review, 34–45.

Bach, V., Møller, J. S., 2003. Correlation at low temperature. I. Exponential decay. J. Funct. Anal. 203 (1), 93–148.

Balinski, M., Laraki, R., 2011. Majority Judgment: Measuring, Ranking, and Electing. MIT Press, Boston, MA.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.

Barthélémy, J.-P., Janowitz, M. F., 1991. A formal theory of consensus. SIAM J. Discrete Math. 4 (3), 305–322.

Barthélémy, J.-P., Monjardet, B., 1981. The median procedure in cluster analysis and social choice theory. Math. Social Sci. 1 (3), 235–267.

Barthélémy, J.-P., Monjardet, B., 1988. The median procedure in data analysis: new results and open problems. In: Classification and related methods of data analysis (Aachen, 1987). North-Holland, Amsterdam, pp. 309–316.

Ben-Yashar, R., Paroush, J., 2001. Optimal decision rules for fixed-size committees in polychotomous choice situations. Soc. Choice Welf. 18 (4), 737–746.

Berend, D., Sapir, L., 2007. Monotonicity in Condorcet's jury theorem with dependent voters. Soc. Choice Welf. 28 (3), 507–528.

Berg, S., 1993a. Condorcet's jury theorem, dependency among jurors. Soc. Choice Welf. 10 (1), 87–95.

Berg, S., 1993b. Condorcet's jury theorem revisited. European Journal of Political Economy 9, 437–446.

Berg, S., 1994. Evaluation of some weighted majority decision rules under dependent voting. Math. Social Sci. 28 (2), 71–83.

Berg, S., 1996. Condorcet's jury theorem and the reliability of majority voting. Group Decision and Negotiation 5, 229–238.

Black, D. S., 1958. Theory of committees and elections. Cambridge UP, Cambridge, UK.

Boland, E. J., 1989. Majority systems and the Condorcet jury theorem. The Statistician 38, 181–189.

Boland, E. J., Proschan, E., Tong, Y., 1989. Modelling dependence in simple and indirect majority systems. Journal of Applied Probability 26, 81–88.

Cervone, D. P., Zwicker, W. S., 2009. Convex decompositions. J. Convex Anal. 16 (2), 367–376.

Clemen, R. T., Winkler, R. L., 1999. Combining probability distributions from experts in risk analysis. Risk analysis 19 (2), 187–203.

Condorcet, M. d., 1785. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris.

DeGroot, M. H., 1974. Reaching a consensus. Journal of the American Statistical Association 69 (345), 118–21.

Dietrich, F., List, C., 2004. A model of jury decisions where all jurors have the same evidence. Synthese 142 (2), 175–202.

Dietrich, F., Spiekermann, K., March 2013a. Epistemic democracy with defensible premises. Economics and Philosophy 29 (01), 87–120.

Dietrich, F., Spiekermann, K., 2013b. Independent opinions? on the causal foundations of belief formation and jury theorems. Mind 122 (487), 655–685.

Downs, A., 1957. An Economic Theory of Democracy. Harper.

Elster, J. (Ed.), 1998. Deliberative Democracy. Cambridge UP.

Estlund, D. M., 1994. Opinion leaders, independence, and condorcet's jury theorem. Theory and Decision 36 (2), 131–162.

Fishkin, J., 2009. When the people speak: Deliberative democracy and public consultation. Oxford University Press, Oxford.

Fishkin, J. S., Laslett, P. (Eds.), 2003. Debating Deliberative Democracy. Blackwell.

Galton, F., March 1907. Vox populi. Nature 75, 450–451.

Genest, C., Zidek, J. V., 1986. Combining probability distributions: a critique and an annotated bibliography. Statist. Sci. 1, 114–148.

Golub, B., Jackson, M. O., February 2010. Naïve learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics 2 (1), 112– 149.

Grofman, B., Owen, G., Feld, S., 1989. Proving a distribution-free generalization of the Condorcet jury theorem. Mathematical Social Sciences 17, 1–16.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux.

Kanazawa, S., 1998. A brief note on a further refinement of the Condorcet jury theorem for heterogeneous groups. Math. Social Sci. 35 (1), 69–73.

Kaniovski, S., 2009. An invariance result for homogeneous juries with correlated votes. Math. Social Sci. 57 (2), 213–222.

Kaniovski, S., 2010. Aggregation of correlated votes and Condorcet's jury theorem. Theory and Decision 69 (3), 453–468.

Kaniovski, S., Zaigraev, A., 2011. Optimal jury design for homogeneous juries with correlated votes. Theory and Decision 71 (4), 439–459.

Karotkin, D., Paroush, J., 2003. Optimum committee size: quality-versus-quantity dilemma. Soc. Choice Welf. 20 (3), 429–441.

Ladha, K. K., 1992. The Condorcet Jury Theorem, free speech, and correlated votes. American Journal of Political Science 36, 617–634.

Ladha, K. K., 1993. Condorcet's jury theorem in light of de Finetti's theorem. Majority-rule voting with correlated votes. Soc. Choice Welf. 10 (1), 69–85.

Ladha, K. K., 1995. Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. Journal of Economic Behaviour and Organization 26, 353–372.

Lam, L., Suen, C. Y., 1996. Majority vote of even and odd experts in a polychotomous choice situation. Theory and Decision 41 (1), 13–36.

Landemore, H., 2013. Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many. Princeton University Press.

Landemore, H., Elster, J., 2012. Collective Wisdom: Principles and Mechanisms. Cambridge University Press.

Leib, E. J., 2004. Deliberative democracy in America: A proposal for a popular branch of government. Pennsylvania State University Press.

List, C., Goodin, R. E., 2001. Epistemic democracy: Generalizing the Condorcet Jury Theorem. Journal of Political Philosophy 9 (3), 277–306.

Martinelli, C., 2006. Would rational voters acquire costly information? Journal of Economic Theory 129 (1), 225–251.

McLennan, A., 1998. Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents. American Political Science Review 92 (02), 413–418.

McMorris, F. R., Mulder, H. M., Powers, R. C., 2000. The median function on median graphs and semilattices. Discrete Appl. Math. 101 (1-3), 221–230.

Nehring, K., Pivato, M., 2017. The median rule in judgement aggregation. (preprint).

Newman, M. E. J., Watts, D. J., Strogatz, S. H., 2002. Random graph models of social networks. Proceedings of the National Academy of Sciences 99 (suppl 1), 2566–2572.

Nitzan, S., Paroush, J., 1984. The significance of independent decisions in uncertain dichotomous choice situations. Theory and Decision 17 (1), 47–60.

Owen, G., 1986. Fair indirect majority rules. In: Grofman, B., Owen, G. (Eds.), Information Pooling and Group Decision Making. Jai Press, Greenwich, CT.

Page, S., 2008. The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. Princeton University Press.

Paroush, J., 1998. Stay away from fair coins: a Condorcet jury theorem. Soc. Choice Welf. 15 (1), 15–20.

Peleg, B., Zamir, S., 2012. Extending the Condorcet jury theorem to a general dependent jury. Soc. Choice Welf. 39 (1), 91–125.

Penrose, O., Lebowitz, J. L., 1974. On the exponential decay of correlation functions. Comm. Math. Phys. 39, 165–184.

Pivato, M., 2013a. Variable-population voting rules. J. Math. Econom. 49 (3), 210–221.

Pivato, M., 2013b. Voting rules as statistical estimators. Soc. Choice Welf. 40 (2), 581–630.

Procacci, A., Scoppola, B., 2001. On decay of correlations for unbounded spin systems with arbitrary boundary conditions. J. Statist. Phys. 105 (3-4), 453–482.

Régnier, S., 1977. Stabilité d'un opérateur de classification. Math. Sci. Humaines 60, 21–30, (Reprinted in Math. Sci. Humaines, Tome 82, pp.75-84 (1983)).

Shapley, L., Grofman, B., 1984. Optimizing group judgemental accuracy in the presence of interdependencies. Public Choice 43, 329–343.

Sunstein, C., 2009. Republic.com 2.0. Princeton University Press.

Sunstein, C. R., 2003. The law of group polarization. In: Fishkin and Laslett (2003), pp. 80–101.

Xia, L., 2015. Generalized decision scoring rules: Statistical, computational, and axiomatic properties. In: Proceedings of the 15th ACM Conference on Electronic Commerce (EC-15). Association for Computing Machinery, Chicago.

Xia, L., Conitzer, V., 2008. Generalized scoring rules and the frequency of coalitional manipulability. In: Proceedings of the 9th ACM Conference on Electronic Commerce (EC-08). Association for Computing Machinery, Chicago, pp. 109–118.

Young, H. P., 1986. Optimal ranking and choice from pairwise decisions. In: Grofman, B., Owen, G. (Eds.), Information pooling and group decision making. JAI press, Greenwich, CT, pp. 113–122.

Young, H. P., 1988. Condorcet's theory of voting. American Political Science Review 82 (4), 1231–1244.

Young, H. P., Winter 1995. Optimal voting rules. Journal of Economic Perspectives 9 (1), 51–64.

Young, H. P., 1997. Group choice and individual judgments. In: Mueller, D. C. (Ed.), Perspectives on public choice: A handbook. Cambridge UP, Cambridge, UK, Ch. 9, pp. 181–200.

Zwicker, W. S., 2008. Consistency without neutrality in voting rules: When is a vote an average? Math. Comput. Modelling 48 (9-10), 1357–1373.