The median rule in judgement aggregation

Nehring, Klaus and Pivato, Marcus

Department of Economics, University of California, Davis, THEMA, Université de Cergy-Pontoise

30 January 2018

The median rule in judgement aggregation*

Klaus Nehring† and Marcus Pivato‡

January 30, 2018

Abstract

A *judgement aggregation rule* takes the views of a collection of voters over a set of interconected issues, and yields a logically consistent collective view. The *median rule* is a judgement aggregation rule that selects the logically consistent view which minimizes the *average distance* to the views of the voters (where the "distance" between two views is the number of issues on which they disagree). In the special case of preference aggregation, this is called the *Kemeny rule*. We show that, under appropriate regularity conditions, the median rule is the unique judgement aggregation rule which satisfies three axioms: *Ensemble Supermajority Efficiency*, *Reinforcement*, and *Continuity*. Our analysis covers aggregation problems in which different issues have different weights, and in which the consistency restrictions on input and output judgments may differ.

JEL classification: D71.
Keywords: Judgement aggregation; majoritarian; reinforcement; consistency; median.

1 Introduction

In judgment aggregation, a group is faced with a joint decision; frequently, the members of the group disagree about which decision the group should take and/or the grounds for the decision. Complex decisions can often be described as an interrelated set of judgments on a set of binary issues subject to some admissibility constraint. Admissibility constraints may be logical, normative or physical.

*Versions of the paper have been presented at the 2016 Meeting of Society for Social Choice and Welfare (Lund), the Workshop on Decision Making and Contest Theory (Kibbutz Ein Gedi, 2016), D-TEA (Paris 2017), and various seminars. We are grateful to the participants at these presentations for their valuable suggestions. We also thank Florian Brandl for helpful comments.

†Department of Economics, UC Davis, California, USA. Email: kdnehring@ucdavis.edu.

‡THEMA, Université de Cergy-Pontoise, France. Email: marcuspivato@gmail.com. An early version of this paper was written while Pivato was at the Department of Mathematics of Trent University in Canada. This research was supported by NSERC grant #262620-2008 and Labex MME-DII (ANR11-LBX-0023-01).

Judgment aggregation theory attempts to determine normative criteria on how to best resolve the disagreement at hand. Such "resolution of disagreement" can be understood in two ways: *compromise* or *consensus*. By "consensus", we mean a well-supported inference from the position of a hypothetical impartial observer. By "compromise" we mean the best accommodation of each member's views from their own perspective. In a slogan, the consensus perspective aims to maximize *plausibility*, while the compromise perspective aims to maximize *concordance.*

Judgment aggregation pertains both to groups that act as sovereigns ("electorates"), as in democratic elections and referenda, and to groups that act as delegate bodies ("committees"), such as multi-member courts, corporate boards, central banks. A key rationale of the existence of many committees is the production of decisions that are sound from an independent third-party perspective; in those cases, the judgment aggregation framework seems especially apt, arguably often more so than the more common and established preference aggregation framework.

Given a profile of views by the group members (henceforth: "voters"), which view should the group adopt? Which view enjoys the highest "support" ("plausibility" or "concordance")? In recent work (Nehring et al., 2014, 2016; Nehring and Pivato, 2014, 2018), we have explored a "majoritarian" approach to this question. Its hallmark is to evaluate support issue by issue in terms of the sign and size of issue-wise majorities. In simple cases in which the issue-wise majorities happen to produce a jointly feasible view, on the majoritarian approach, this view enjoys the highest support, and should thus be adopted by the group. But issue-wise majorities may well not yield a consistent view. In the context of preference aggregation, this is the well-known Condorcet paradox. Analoguous inconsistencies are very common in judgment aggregation (Guilbaud, 1952; Kornhauser and Sager, 1986; List and Pettit, 2002), and have stimulated an intensive investigation in economic theory, philosophy, and computer science; see e.g. List and Puppe (2009) and Mongin (2012) for recent surveys.

A satisfactory normative account thus needs to be able to deal with the Condorcet-inconsistent cases. In contrast to the focus on impossibility results in much of the literature just referenced, in this paper we make a case for a particular normative decision criterion: the *median rule.* The median rule maximizes the total numerical support (number of votes) for a view, summed over all issues. (Generically, this maximizer is unique.) Equivalently, the median rule minimizes the average distance to the views of the voters (where the "distance" between two views is measured by the number of issues on which they differ).

Our characterization of the median rule is based on three normative axioms. First, majoritarianism over multiple issues is encoded in an axiom called "Ensemble Supermajority Efficiency" (ESME), which itself is an extension of the normatively more basic principle of "Supermajority Efficiency" (SME). The SME principle says that if, in the comparison of two admissible views $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}$ agrees with the majority on more issues than $\mathbf{y}$, then $\mathbf{y}$ is inferior as a group view, and thus should not be adopted by the group. Furthermore, and more demandingly, for any fixed 'quorum' q, we count how many supermajorities of size at least q agree with $\mathbf{x}$ or with $\mathbf{y}$. If $\mathbf{x}$ agrees with at least as many size-q supermajorities as $\mathbf{y}$ for all values of q, and $\mathbf{x}$ agrees with at *more* size-q supermajorities as $\mathbf{y}$ for *some* value of

q, then again $\mathbf{y}$ is inferior as a group view, and should not be adopted by the group. Thus, SME takes into account the size of the supporting supermajorities in a manner analogous to first-order stochastic dominance in decision theory. The ESME axiom extends SME to "ensembles" of judgment aggregation problems consisting of multiple instances of same type of judgment aggregation problem but with different profiles.

Second, the axiom of Reinforcement says that if two subpopulations independently choose the same view under the rule, then the combined population should also choose this view under the rule. It is a standard, highly versatile axiom originally due to Smith (1973) and Young (1974). Finally, Continuity asserts that the group judgment is robust under small perturbations of the distribution of input judgments. Our first main result, Theorem 1, characterizes the median rule as the unique judgment aggregation rule satisfying ESME, Reinforcement and Continuity. Theorem 1 is based on a weak minimal richness on the input space called "thickness".

Theorem 1 treats all issues symmetrically. This is warranted in many standard applications whose structure is sufficiently symmetric such as the aggregation of rankings (linear or weak orderings), of classifiers (equivalence relations) or in multi-winner choice problems. But other applications lack these symmetries; furthermore, different issues, whether or not formally symmetric, may be given different " importance". For example, consider truth-functional aggregation, which was the focus of much of the early literature in judgment aggregation inspired by Kornhauser and Sager's (1986) "doctrinal paradox". In truth-functional aggregation, one or more "conclusion judgments" are logically ("truth-functionally") determined by a number of "premise judgments". Condorcet inconsistency takes the form of the "discursive dilemma": issuewise aggregation of majorities on the premises may well determine (by truth-functional implication) judgments on the conclusions that differ from the majority judgment on these conclusions. The discursive dilemma can be resolved via the median rule by trading off majority overrides on the premises against majority overrides on the conclusions. However, in view the structural and conceptual asymmetry between premises and conclusions, they have different standing, and it would appear quite arbitrary to give them equal weight.

In section 5, we thus generalize the analysis to *weighted* judgment contexts in which different issues have different weights. The definitions of SME, ESME and the median rule generalize naturally. However, as shown by counterexample, Theorem 1 does not carry over in full generality. The characterization for weighted judgement contexts in Theorem 2 must invoke not just richness conditions on the space of admissible *input* judgments, but also restrictions on the "combinatorial geomety" of the space of admissible *output* judgments. We note that, while previous work on judgment aggregation assumed the input and output spaces to be the same, we allow them to differ. This adds useful additional generality at very modest cost in execution.

Due to the abstraction and generality of our judgment aggregation framework, it has a broad and diverse range of applications. We thus illustrate our concepts and results in a number of examples, including applications to approval voting on committees with composition constraints, assignment problems, uniform treatment of heterogeneous cases, missing information, and multiple criteria. See Nehring and Puppe (2007), Nehring and

Pivato (2011) or Nehring et al. (2014) for many more examples.

The main comparison result in the literature is the remarkable characterization of the median rule in the aggregation of linear orderings ("rankings") by Young and Levenglick (1978). In this setting (mathematically equivalent to the setting of Arrovian preference aggregation), the median rule is also known as the Kemeny (1959) rule.

For the problem of aggregating ordinal rankings, our assumptions are broadly similar philosophically to those of Young and Levenglick (1978), but differ in the specifics, our assumptions being on the whole 'stronger'. We thus do not claim that when applied to the aggregation of rankings, our Theorem 1 improves on Young and Levenglick's result. Our aims are simply different. While their result relies heavily on particular features of the combinatorial geometry of the space of ordinal rankings, Theorem 1 is a "one-size-fits-all" result that covers a wide range of judgment aggregation problems. (See the end of section 4 for a more detailed comparison)

This paper belongs to a larger project exploring multi-issue majoritarianism in judgment aggregation. In particular, the main results in the present paper rely on results in a companion paper (Nehring and Pivato, 2018), which show that judgment aggregation rules that satisfy ESME and Continuity are representable as "additive majority rules". Additive majority rules can be viewed as non-linear generalizations of the median rule; they evaluate views not simply in terms of the (weighted) sum of numeric issue-wise majorities, but in terms of the (weighted) sum of issue-wise majority *gains*, which are possibly non-linear transformations of these numeric issue-wise majorities. Theorems 1 and 2 can be read as showing that Reinforcement implies the existence of a representation as an additive majority rule with a linear gain function. While plausible, the asserted connection between Reinforcement and linearity is less straightforward than it may look, and the proof encounters a number of non-trivial obstacles, among them the need for substantive structural assumptions on the judgment space, which play no role in obtaining an additive majority representation in the first place.

The rest of this paper is organized as follows. Section 2 sets up the formal framework. Section 3 introduces the axioms of ESME and Continuity, and explains that additive majority rules are the only rules satisfying these axioms. Section 4 states our axiomatic characterization of the median rule for unweighted judgement aggregation contexts. Section 5 introduces *weighted* judgement contexts, and Section 6 extends our axiomatic characterization to such contexts. All proofs are in the Appendices.

2 Judgement aggregation

Let $\mathcal{K}$ be a finite set of logical propositions or *issues*, each of which can be either affirmed or denied. A *view* is an assignment of an assertoric (Yes-No) value to each issue, represented by an element of $\{\pm 1\}^{\mathcal{K}}$. A *judgement space* is a collection of views —that is, a subset of $\{\pm 1\}^{\mathcal{K}}$ —determined by certain constraints. These constraints can arise in several ways: as a matter of logical consistency (as in truth-functional aggregation problems), as a matter of "rational coherence" (as in transitivity conditions on orderings) or as mere "feasibility" (as in multi-winner choice problems).

Example 1. (*Aggregation of rankings*) Let $\mathcal{A}$ be a finite set of alternatives. We can represent the set of all strict ordinal rankings of $\mathcal{A}$ as a judgement space $\mathcal{X}_{\mathcal{A}}^{\mathrm{rk}} \subset \{\pm 1\}^{\mathcal{K}}$, where the elements of $\mathcal{K}$ represent assertions of the form "$a \succ b$" for each pair $a, b \in \mathcal{A}$, and admissibility is given by transitivity.

To be technically specific, let $\mathcal{K} \subset \mathcal{A} \times \mathcal{A}$ be a subset such that $(a, a) \notin \mathcal{K}$ for any $a \in \mathcal{A}$, and for each distinct $a, b \in \mathcal{A}$, exactly *one* of the pairs (a, b) or (b, a) is in $\mathcal{K}$. Any complete, antisymmetric binary relation $\succ$ on $\mathcal{A}$ can be represented by a unique element $\mathbf{x}$ of $\{\pm 1\}^{\mathcal{K}}$ by setting $x_{ab} = 1$ if $a \succ b$, whereas $x_{ab} = -1$ if and only if $a \prec b$. Let $\mathcal{X}_{\mathcal{A}}^{\mathrm{rk}}$ be the set of all elements of $\{\pm 1\}^{\mathcal{K}}$ corresponding to *ordinal rankings* of $\mathcal{A}$.[1] Judgement aggregation on the space of ordinal rankings $\mathcal{X}_{\mathcal{A}}^{\mathrm{rk}}$ is thus formally equivalent to classical Arrovian preference aggregation. ♢

Example 2. (*Classifier aggregation*) Likewise, we can represent the set of all equivalence relations on $\mathcal{A}$ as a judgement space $\mathcal{X}_{\mathcal{A}}^{\mathrm{eq}} \subset \{\pm 1\}^{\mathcal{J}}$, where the elements of $\mathcal{J}$ represent assertions of the form "$a \approx b$", for each pair $a, b \in \mathcal{A}$. Again, to be formally specific, let $\mathcal{K}$ be the set of all two-element subsets of $\mathcal{A}$. Any symmetric, reflexive binary relation $\sim$ on $\mathcal{A}$ can be represented by a unique element $\mathbf{x}$ of $\{\pm 1\}^{\mathcal{K}}$ by setting $x_{ab} = 1$ if and only if $a \sim b$. Let $\mathcal{X}_{\mathcal{A}}^{\mathrm{eq}}$ be the set of all elements of $\{\pm 1\}^{\mathcal{K}}$ corresponding to *equivalence relations* on $\mathcal{A}$.[2] Judgement aggregation on $\mathcal{X}_{\mathcal{A}}^{\mathrm{eq}}$ arises when each voter has her own way of classifying the elements of $\mathcal{A}$ into equivalence classes, and the group must agree on some common classification system. ♢

Other judgement spaces represent common collective decision problems such as resource allocation, committee selection, or taxonomic classification. One particularly well-known class of examples are the so-called *truth-functional* aggregation problems. In this case, the issues in $\mathcal{K}$ are divided into two classes: "premises" and "conclusions", and the truth-values of the conclusions are logically entailed by the truth-values of the premises. The space $\mathcal{X}$ is then the set of all logically consistent assignments of truth values to the premises and conclusions. See Nehring and Puppe (2007), Nehring and Pivato (2011) or Nehring et al. (2014) for many more examples.

Judgment aggregation rules map profiles of views to a group view or set of views. Typically, both outputs and inputs are subject to feasibility or logical consistency constraints, which are encoded by two judgement spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. In many cases, the restrictions on inputs and outputs are the same (so that $\mathcal{X} = \mathcal{Y}$), but they need not be. For example, one might require output views to be fully rationally coherent (e.g. transitive), but allow input views that are not, for example to accommodate bounded rationality in voters. Or output views may take into account feasibility consideration, while input views

[1] Formally: $\mathcal{X}_{\mathcal{A}}^{\mathrm{rk}}$ is the set of all $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$ satisfying the following transitivity constraint, for all distinct $a, b, c \in \mathcal{A}$: if ($x_{ab} = 1$ or $x_{ba} = -1$), and ($x_{bc} = 1$ or $x_{cb} = -1$), then ($x_{ac} = 1$ or $x_{ca} = -1$).

[2] Formally: $\mathcal{X}_{\mathcal{A}}^{\mathrm{eq}}$ is the set of all $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$ satisfying the transitivity constraint $(x_{ab} = x_{bc} = 1) \Longrightarrow (x_{ac} = 1)$, for all distinct $a, b, c \in \mathcal{A}$.

do not. (As in single- or multi-winner Approval voting). We thus define an (unweighted) ***judgement context*** to be a triple $\boldsymbol{\mathcal{C}} := (\mathcal{K}, \mathcal{X}, \mathcal{Y})$, where $\mathcal{K}$ is a (finite) set of issues and $\mathcal{X} \subseteq \mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}}$, with $\mathcal{X}$ being the "input space", and $\mathcal{Y}$ being the "output space". We assume that $\mathcal{X} \subseteq \mathcal{Y}$; i.e. output views must satisfy all admissibility restrictions that input views do, and maybe more. A ***profile*** is a function $\mu : \mathcal{Y} \longrightarrow [0,1]$ such that $\sum_{\mathbf{y} \in \mathcal{Y}} \mu(\mathbf{y}) = 1$. This represents a population of weighted voters; for each $\mathbf{y} \in \mathcal{Y}$, $\mu(\mathbf{y})$ is the total weight of the voters who hold the view $\mathbf{y}$. If all voters have the same weight, then $\mu(\mathbf{y})$ is simply the *proportion* of the electorate which holds the view $\mathbf{y}$. But we allow the possibility that different voters have different weights, e.g. because of different levels of expertise or different stakes in the outcome. By summarizing the voters' views with a function $\mu : \mathcal{Y} \longrightarrow [0,1]$, we abstract from the exact number of voters, and we render them anonymous, except for their weights: voters with the same weight are indistinguishible in our model.

If $\mathcal{Y}$ is a judgement space, then we define $\Delta(\mathcal{Y})$ to be the set of all profiles on $\mathcal{Y}$. If $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \mathcal{X}, \mathcal{Y})$ is a judgement context, then we define $\Delta(\boldsymbol{\mathcal{C}}) := \Delta(\mathcal{Y})$. A ***judgement problem*** is an ordered pair $(\boldsymbol{\mathcal{C}}, \mu)$, where $\boldsymbol{\mathcal{C}}$ is a judgement context, and $\mu \in \Delta(\boldsymbol{\mathcal{C}})$. *Judgement aggregation* is the process of converting such a judgement problem into a view (or set of views) in $\mathcal{X}$. A ***judgement aggregation rule*** on $\boldsymbol{\mathcal{C}}$ is a correspondence $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$, which produces a nonempty (usually singleton) subset $F(\mu) \subseteq \mathcal{X}$ for any profile $\mu \in \Delta(\mathcal{Y})$. (Sometimes, we will write "$F(\mathcal{X}, \mu)$" instead of "$F(\mu)$").

The *median rule* is a particularly attractive judgement aggregation rule. To define it, we need some notation. Recall that $\mathcal{X} \subseteq \mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}} \subset \mathbb{R}^{\mathcal{K}}$. Thus, each view $\mathbf{y} \in \mathcal{Y}$ can be regarded as a vector in $\mathbb{R}^{\mathcal{K}}$. For any profile $\mu \in \Delta(\mathcal{Y})$, we define its ***majority vector***

$$\widetilde{\boldsymbol{\mu}} \quad := \quad \sum_{\mathbf{y} \in \mathcal{Y}} \mu(\mathbf{y})\, \mathbf{y} \;\in\; [-1,1]^{\mathcal{K}}. \tag{1}$$

For all $k \in \mathcal{K}$, we have $\widetilde{\mu}_k > 0$ if a (weighted) majority of voters affirm or support the issue k, whereas $\widetilde{\mu}_k < 0$ if a majority deny or oppose k. The ***majority ideal*** is the element $\mathbf{x}^{\mu} \in \{\pm 1\}^{\mathcal{K}}$ defined by setting $x_k^{\mu} := \operatorname{sign}(\widetilde{\mu}_k)$ for all $k \in \mathcal{K}$.[3] However, for many profiles $\mu \in \Delta(\mathcal{Y})$, it turns out that $\mathbf{x}^{\mu} \notin \mathcal{X}$. (This can happen even when $\mathcal{Y} = \mathcal{X}$.) In other words, it is frequently impossible to agree with the μ-majority in *every* issue in $\mathcal{K}$, while respecting the underlying logical constraints which define the judgement space $\mathcal{X}$.

Informally, the median rule maximizes the *average agreement* with μ-majorities across all the issues in $\mathcal{K}$. Formally, for all $\mu \in \Delta(\mathcal{Y})$, we define

$$\operatorname{Median}(\mathcal{X}, \mu) \quad := \quad \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \sum_{k \in \mathcal{K}} x_k\, \widetilde{\mu}_k. \tag{2}$$

For any $\mathbf{x} \in \mathcal{X}$, let $\mathbf{x} \bullet \widetilde{\boldsymbol{\mu}} := \sum_{k \in \mathcal{K}} x_k\, \widetilde{\mu}_k$. Then we can rewrite (2) in a simpler form:

$$\operatorname{Median}(\mathcal{X}, \mu) \quad := \quad \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \mathbf{x} \bullet \widetilde{\boldsymbol{\mu}}, \qquad \text{for all } \mu \in \Delta(\mathcal{Y}). \tag{3}$$

As we noted in the introduction, in the special case of the aggregation of rankings (i.e. when $\mathcal{X} = \mathcal{Y} = \mathcal{X}_{\mathcal{A}}^{\mathrm{rk}}$), the median rule is equivalent to the Kemeny rule.

[3] For simplicity, we assume in this paragraph that $\widetilde{\mu}_k \neq 0$ for all $k \in \mathcal{K}$; this is not essential.

There is another way to define and motivate the median rule via a natural notion of distance due to Kemeny (1959); see Monjardet (2008) for a broad survey. For any $\mathbf{x}, \mathbf{y} \in \{\pm 1\}^{\mathcal{K}}$, we define their *Hamming distance* by $d(\mathbf{x}, \mathbf{y}) := \#\{k \in \mathcal{K};\ \ x_k \neq y_k\}$. It is easy to see that the median rule selects the view(s) in $\mathcal{X}$ that minimize the *average Hamming distance* to the views of the voters; in the terminology of Miller and Osherson (2009) and Lang et al. (2011), it is a *distance-minimizing* rule.[4] Distance-minimizing rules are themselves a subclass of *scoring rules*, which have been studied by Dietrich (2014).[5]

This metric interpretation is particularly appealing when the task of judgment aggregation is to find an optimal compromise. From this perspective, the Hamming distance of the collective view from the view of any voter is a natural measure of the "burden of compromise" imposed on that voter, and the median view(s) are those that minimize the aggregate burden of compromise.

3 Additive majority rules

Median rules are a special case of "additive majority rules" (AMRs), and the proofs of our main results in Sections 4 and 5 rely on broader results for AMRs obtained in the companion paper (Nehring and Pivato, 2018). Like the median rule, *additive majority rules* try to maximize the "total agreement with majorities", where the "total" is taken by summing over all issues in $\mathcal{K}$, and where the "agreement with majorities" is measured by applying an increasing function (called the *gain function*) to the coordinates of the majority vector $\widetilde{\boldsymbol{\mu}}$. This allows, in particular, larger majorities (especially unanimous or almost unanimous majorites) to carry a disproportionately greater weight than smaller majorities. The added generality appears potentially useful especially from a consensus perspective, from which non-linearities in the gain function can naturally be interpreted as reflecting non-linearities in the plausibility of (evidential support for) judgments as a function of the balance of majorities supporting them.

In some cases, one might even want larger majorities to lexicographically dominate smaller majorities. To allow for such possibilities, we must allow the gain function to take infinite and/or infinitesimal values. Formally, this can be done letting the gain function take values in some linearly ordered field containing, and possibly strictly including the reals. By the results of Nehring and Pivato (2018), the hyperreal numbers are sufficiently general; thus, we define the codomain of the gain function as the linearly ordered field of *hyperreal numbers* ${}^*\mathbb{R}$,[6] and let $\phi : [-1, 1] \longrightarrow {}^*\mathbb{R}$ be an increasing function (the *gain function*). The *additive majority rule* $F_\phi : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is defined:

$$\text{For all } \mu \in \Delta(\mathcal{Y}), \qquad F_\phi(\mu) \quad := \quad \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmax}} \left(\sum_{k \in \mathcal{K}} \phi(x_k \, \widetilde{\mu}_k) \right). \tag{4}$$

[4] Miller and Osherson (2009) call it *Prototype*, while Lang et al. (2011) call it $R^{d_H, \Sigma}$.

[5] In Dietrich's terminology, the median rule is the *simple* scoring rule.

[6] A reader who feels on unfamiliar territory is invited to focus on the special case of real-valued gain functions. All that is needed for the present paper (including proofs) is the elementary arithmetic for addition and multiplication of linearly ordered fields, which is exactly the same as that for the real numbers.

In particular, the median rule is an additive majority rule. To see this, let $\phi(x) := x$ for all $x \in [-1, 1]$; then formula (4) reduces to the formula (2). The main result of this section stated below is adapted from Nehring and Pivato (2018), and is a key step in our axiomatic characterization of the median rule. It states that additive majority rules are the *only* judgement aggregation rules that can simultaneously satisfy two axioms: Ensemble Supermajority Efficiency and Continuity.

Supermajority efficiency. For any $\mathbf{x} \in \mathcal{X}$ and $q \in [0, 1]$, we define $\gamma^{\mu}_{\mathbf{x}}(q) := \#\{k \in \mathcal{K};\ x_k\, \widetilde{\mu}_k \geq q\}$; this yields a non-increasing function $\gamma^{\mu}_{\mathbf{x}} : [0, 1] \longrightarrow \mathbb{R}$. We say $\mathbf{x} \in \mathcal{X}$ is *supermajority efficient* (SME) for the judgement problem $(\mathcal{X}, \mu)$ if there does not exist any $\mathbf{z} \in \mathcal{X}$ such that $\gamma^{\mu}_{\mathbf{z}}(q) \geq \gamma^{\mu}_{\mathbf{x}}(q)$ for all $q \in [0, 1]$, with strict inequality for some q. Let $\mathrm{SME}\,(\mathcal{X}, \mu)$ be the set of such views. If $\mathbf{x} \in \mathrm{SME}\,(\mathcal{X}, \mu)$, then it is impossible to change a coordinate of $\mathbf{x}$ to capture one more μ-supermajority of size q, without either losing at least one μ-supermajority of size $q' \geq q$, or losing at least two μ-supermajorities of size $q' \leq q$. In the Condorcet consistent case, i.e. if the majority ideal $\mathbf{x}^{\mu}$ is in $\mathcal{X}$, then $\mathrm{SME}\,(\mathcal{X}, \mu) = \{\mathbf{x}^{\mu}\}$. We will say that a judgement aggregation rule $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is *supermajority efficient* (SME) if $F(\mu) \subseteq \mathrm{SME}\,(\mathcal{X}, \mu)$ for any choice of $\mu \in \Delta(\mathcal{Y})$. It is easily verified that any additive majority rule is SME (Nehring and Pivato, 2018). The next example illustrates the significance of supermajority efficiency.

Example 3. (*Voting on Committees*) A committee of L members is to be chosen. For $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$, let $\#\mathbf{x} := \#\{k \in \mathcal{K};\ x_k = 1\}$. With $\mathcal{K}$ denoting the set of candidates, the set of feasible committees can be written as the set $\mathcal{X}^{\mathcal{K}}_L = \{\mathbf{x} \in \{\pm 1\}^{\mathcal{K}} : \#\mathbf{x} = L\}$. This encoding breaks down the selection of a committee into $|\mathcal{K}|$ binary issues, each of the form "should candidate k be a member of the committee?"

There are two natural input spaces. On the one hand, with $\mathcal{Y} = \mathcal{X}$, each voter's input consists of a feasible committee, naturally interpreted as her own view of the best committee. On the other hand, with $\mathcal{Y} = \{\pm 1\}^{\mathcal{K}}$, a voter's input consists of an independent judgment on each candidate; this could be interpreted as a judgment of "competency" or "merit" of the candidate. In line with the literature, we will refer to such judgments as judgments of "approval", and the domain $\{\pm 1\}^{\mathcal{K}}$ as the *approval domain* on $\mathcal{K}$.

Consider any profile $\mu \in \Delta\,(\mathcal{Y})$; to sidestep ties, assume for simplicity that all majority margins $\widetilde{\mu}_k$ are different. In such cases, the SME criterion singles out a unique committee as optimal, namely the committee composed of the L candidates with the highest majority margins $\widetilde{\mu}_k$ (regardless of whether or not these are positive). Since AMRs are supermajority efficient , this agrees with the output of any AMR, including the median rule. With the approval domain and $L = 1$, SME thus selects the candidate k with the highest $\widetilde{\mu}_k$; this is just the approval voting rule of Brams and Fishburn (1983). With $L > 1$, SME yields "multi-winner approval voting". Multi-winner approval voting is rationalized here via SME which evaluates views (i.e. committees) issue by issue (i.e. candidate by candidate).

Multi- winner approval voting evaluates committees exclusively based on the "stand-alone" evaluation of by their judged merit. In many situations, there is also a shared interest in the composition of the committee, for example to ensure broad expertise or broad representation of perspectives or stake-holders. A simple way to incorporate such considerations is by way of exogenous *composition constraints* which restrict the set of admissible committees $\mathcal{X}$ to a strict subset of $\mathcal{X}_L^{\mathcal{K}}$.

For example, in the selection of a university-wide committee, one may want to ensure an adequate representation of types of disciplines $j \in \mathcal{J}$ (such as humanities, social sciences, natural sciences, engineering) by imposing a minimum size of the membership of each group L_j. With $\{\mathcal{K}_j\}_{j\in\mathcal{J}}$ describing the partition of candidates according to their disciplinary type, this composition constraint yields an output space $\mathcal{X} = \{\mathbf{x} \in \mathcal{X}_L^{\mathcal{K}} : \#\{k \in \mathcal{K}_j : x_k = 1\} \geq L_j \text{ for all } j \in \mathcal{J}\}$. Again, except for possible ties, the SME criterion yields unique optimal committees: first, for each group $j \in \mathcal{J}$, select the L_j candidates with the highest majority support in that group; second, fill the remaining $L - \sum_{j\in\mathcal{J}} L_j$ candidates with those among the not-yet selected candidates with the highest majority support across all disciplines.

In more complex situations, more than one type of classification may be considered relevant such as academic rank, gender, or ethnicity, leading to additional, overlapping composition requirements. In those cases, SME might no longer select a uniquely optimal committee; that selection might now depend on the particular AMR used. The median rule in particular selects those comittees x with the highest overall approval $\sum_{k:x_k=1} \widetilde{\mu}_k$; it thus yields a natural formulation of *multi-winner approval voting under composition constraints.* ♢

Ensemble supermajority efficiency. In applying the SME criterion to judgment aggregation *rules*, not just single profiles, one can get additional leverage by considering *ensembles* of judgment problems. Such ensembles consist of N instances of the same judgment context with potentially different profiles $\mu_1, \ldots, \mu_N$. To picture such ensembles concretely, an academic electorate may need to simultaneously appoint committees with different tasks but the same structure (composition constraint). Or, in a sequential version of the same idea, it may need to annually elect a committe for a given task and structure, but potentially different candidates. From a normative standpoint, it suffices to assume that these ensembles are meaningful as hypotheticals; they do not need to be actual features of the judgment problem at hand.[7]

To apply SME formally to ensembles, one needs to represent ensembles of judgment contexts as judgment contexts on their own. To do so, simply enhance the issue space to the $N-$fold disjoint union of (copies of) $\mathcal{K}$ written as $N \cdot \mathcal{K}$, and let the output and input spaces be N-fold Cartesian powers $\mathcal{X}^N := \mathcal{X} \times \cdots \times \mathcal{X}$ and $\mathcal{Y}^N := \mathcal{Y} \times \cdots \times \mathcal{Y}$. Thus, we obtain a new judgment context $\boldsymbol{\mathcal{C}}^N := (N \cdot \mathcal{K}, \mathcal{X}^N, \mathcal{Y}^N)$.

Given a profile $\mu \in \Delta(\mathcal{Y}^N)$, we define its nth marginal $\mu^n \in \Delta(\mathcal{Y})$ to be the profile such that, for any $\mathbf{x} \in \mathcal{Y}$,

$$\mu^n(\mathbf{x}) \quad := \sum_{\substack{(\mathbf{y}_1,\ldots,\mathbf{y}_N)\in\mathcal{Y}^N \\ \text{with } \mathbf{y}_n=\mathbf{x}}} \mu(\mathbf{y}_1, \ldots, \mathbf{y}_N). \tag{5}$$

[7] Their role is thus quite similar to the role of counterfactual profiles in multi-profile restrictions.

An N-tuple of profiles $\{\mu_1, \ldots, \mu_N\}$ can be represented by any profile $\mu \in \Delta\left(\mathcal{Y}^N\right)$ such that, for each $n \in [1 \ldots N]$, $\mu^n = \mu_n$. Consider thus an "ensemble problem" $(\mathcal{C}^N, \mu)$. Denote the normative output for this problem by $F^N(\mu)$. It seems plausible that this normative output should be obtained by applying the rule F factor by factor. In other words, $F^N(\mu) = F(\mu^1) \times \cdots \times F\left(\mu^N\right)$, for all $\mu \in \Delta\left(\mathcal{Y}^N\right)$. Since $(\mathcal{C}^N, \mu)$ is itself a well-defined judgment aggregation problem, SME applies to it, just as it does to 'ordinary' judgment aggregation problems. We are thus lead to the axiom of Ensemble Supermajority Efficiency.

ESME. For any number of instances $N \in \mathbb{N}$ and profile $\mu \in \Delta\left(\mathcal{Y}^N\right)$, any element of $F(\mu^1) \times \cdots \times F\left(\mu^N\right)$ is SME in the judgment aggregation problem $(\mathcal{C}^N, \mu)$.[8]

This axiom requires F to be supermajority efficient. But it requires more, because an ensemble view $\mathbf{x} \in \mathcal{X}^N$ that is SME instance-by-instance need not be SME overall. Indeed, we shall soon see that any aggregation rule that satisfies ESME must be an additive majority rule, as soon as it is continuous in the following sense .

Continuity. For every profile $\mu \in \Delta(\mathcal{Y})$, and every sequence $\{\mu_n\}_{n=1}^{\infty} \subset \Delta(\mathcal{Y})$ with $\lim_{n\to\infty} \mu_n = \mu$, if $\mathbf{x} \in F(\mu_n)$ for all $n \in \mathbb{N}$, then $\mathbf{x} \in F(\mu)$.

This axiom says that the correspondence F is upper hemicontinuous with respect to the usual, Euclidean topology on $\Delta(\mathcal{Y})$. This means that F is robust against small perturbations or errors in the specification of μ. It also means that, if a very large population of voters is mixed with a much smaller population, then the views of the large population essentially determine the outcome of the rule. The following result is an adaptation of a main result in the companion paper Nehring and Pivato (2018).

Proposition 1 *Let F be a judgement aggregation rule on a judgement context $\mathcal{C}$. If F satisfies* ESME *and* Continuity *on $\Delta(\mathcal{C})$, then F is an additive majority rule.*

4 Axiomatic characterization of the median rule

To characterize the median rule, we will need one more axiom in addition to the two which appeared in Proposition 1: *Reinforcement.*

Reinforcement. Let $\mu_1, \mu_2 \in \Delta(\mathcal{Y})$ be two profiles, describing two subpopulations of size S_1 and S_2. Let $c_1 = S_1/(S_1 + S_2)$ and $c_2 = S_2/(S_1 + S_2)$. Then $\mu = c_1\mu_1 + c_2\mu_2$ is the profile of the combined population. If each subpopulation separately endorses some view $\mathbf{x} \in \mathcal{X}$, then the combined population presumably should also endorse this view. The next axiom formalizes this desideratum.

[8] The profile μ also contains information about the joint distribution of views over the different instances of $\mathcal{Y}$. But this extra information is immaterial to ESME, since it is used in neither $F^N\left(\mathcal{X}^N, \mu\right)$ nor $\mathrm{SME}\left(\mathcal{X}^N, \mu\right)$.

Reinforcement. For any profiles $\mu_1, \mu_2 \in \Delta(\mathcal{Y})$ with $F(\mu_1) \cap F(\mu_2) \neq \emptyset$, and any $c_1, c_2 \in (0,1)$ with $c_1 + c_2 = 1$, if $\mu = c_1\,\mu_1 + c_2\,\mu_2$, then $F(\mu) = F(\mu_1) \cap F(\mu_2)$. In other words, for any view $\mathbf{x} \in \mathcal{X}$, we have $\mathbf{x} \in F(\mu_1) \cap F(\mu_2)$ if and only if $\mathbf{x} \in F(\mu)$.

In the present setting, Reinforcement is appealing especially from a compromise perspective. If a particular view $\mathbf{x}$ minimize the aggregate "burden of compromise" within some subpopulation μ_1 and the same view $\mathbf{x}$ happens to minimize the aggregate "burden of compromise" within a disjoint subpopulation μ_2, it stands to reason that $\mathbf{x}$ *ipso facto* minimizes the aggregate "burden of compromise" within the combined population μ. This argument for Reinforcement parallels a standard argument for Reinforcement as an Extended Pareto condition in preference aggregation (Dhillon and Mertens, 1999).

Reinforcement seems less compelling prima facie from the consensus perspective which treats the input judgments as the 'evidential basis' for an outside observer; in particular, the consensus perspective bars an 'extended Pareto' argument for Reinforcement. For example, consider a situation in which there is unanimous agreement on some issue k in one subpopulation with profile μ while there is a near tie in the other subpopulation μ'. At both profiles, the same view $\mathbf{x}$ happens to be selected as 'most plausible' according to F. In the combined population, say $\mu'' = \frac{1}{2}\mu + \frac{1}{2}\mu'$, there is a clear majority on k, but it is far from unanimity. So the profile μ'' is materially distinct as evidence from either μ or μ', and it may very well be sensible to select a view $\mathbf{y}$ different from $\mathbf{x}$ as 'most plausible' given the evidence μ''. This may well happen, for example, at some profiles under an additive majority rule with a non-linear gain function ϕ.

While the case for a linear gain function may not be as compelling from a consensus perspective as it is from a compromise perspective (via Reinforcement), there is still a good case to be made on the basis of a "default principle" of sorts. The next axiom serves as an axiomatic expression of such a default principle. In contrast to Reinforcement, which is a variable-population axiom, the next axiom compares profiles within a fixed population of voters. In a nutshell, it considers how the rule F should respond to a change of opinion in one sub-population while the opinion of the complementary sub-population remains fixed. It says that F should always respond to a given opinion change in the same way, independent of the opinion of the complementary sub-population; it is conceptually analogous to the axiom of *Tradeoff Consistency* in decision theory.

Judgement Consistency. For any $c_1, c_2 \in (0,1)$ with $c_1 + c_2 = 1$, and any profiles $\mu, \mu', \nu, \nu' \in \Delta(\mathcal{Y})$, and any views $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if (a) $\mathbf{y} \in F(c_1\,\mu + c_2\,\nu)$, and (b) $\mathbf{x} \in F(c_1\,\mu' + c_2\,\nu)$, and (c) $\mathbf{y} \notin F(c_1\,\mu' + c_2\,\nu)$, and (d) $\mathbf{x} \in F(c_1\,\mu + c_2\,\nu')$, then (e) $\mathbf{y} \notin F(c_1\,\mu' + c_2\,\nu')$,

The idea here is that F should only select the views in $\mathcal{X}$ which have the greatest 'plausibility' in light of the input judgments. In the above axiom, μ, μ' are profiles describing two possible distributions opinions for a subpopulation S_1 making up a proportion c_1 of the total population, while ν, ν' are profiles describing two possible distributions of opinions for the complementary subpopulation S_2 (making up the proportion $c_2 = 1 - c_1$ of the total population). Hypotheses (a), (b), and (c) say that the shift in the distribution of opinions from μ to μ' shifts the balance of plausbility from $\mathbf{y}$ to $\mathbf{x}$, when the S_2 subpopulation has profile ν. Thus, if we start with another profile $(c_1\,\mu + c_2\,\nu')$ where $\mathbf{x}$ is *already* weakly more plausible than $\mathbf{y}$ (hypothesis (d)), then the same shift of opinion from μ to μ' among subpopulation S_1 should *again* make the $\mathbf{x}$ strictly more plausible than $\mathbf{y}$ (conclusion (e)).

It is easy to verify that the median rule satisfies Judgement Consistency. The next result describes the logical relationship between the last three axioms.

Proposition 2 *If an aggregation rule satisfies* Continuity *and* Judgement Consistency*, then it satisfies* Reinforcement.[9]

To obtain the desired axiomatization of the median rule, we will need a weak structural condition on $\mathcal{Y}$. Recall that $\mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}} \subset \mathbb{R}^{\mathcal{K}}$. Let $\operatorname{conv}(\mathcal{Y})$ be the convex hull of $\mathcal{Y}$ in $\mathbb{R}^{\mathcal{K}}$. We will say that $\mathcal{Y}$ is *thick* if $\operatorname{conv}(\mathcal{Y})$ has dimension $|\mathcal{K}|$. To motivate this assumption, note that $\operatorname{conv}(\mathcal{Y})$ is the set of *majority vectors* associated with admissible profiles μ; that is, $\operatorname{conv}(\mathcal{Y}) = \{\widetilde{\boldsymbol{\mu}} : \mu \in \Delta(\mathcal{Y})\}$. Thus, thickness just says that the domain of profiles is sufficiently "rich".

Most interesting judgment aggregation spaces are thick. For examples, spaces of rankings (Example 1) are thick, as are spaces of classifiers (Example 2). Evidently, the "approval spaces" $\{\pm 1\}^{\mathcal{K}}$ involved in the Committee Voting examples are thick. See Nehring and Pivato (2011) for many other examples. Here is our first main result.

Theorem 1 *Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \mathcal{X}, \mathcal{Y})$ be a judgement context where $\mathcal{Y}$ is thick. Let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be a judgement aggregation rule. Then F satisfies* ESME*,* Continuity*, and* Reinforcement *if and only if F is the median rule.*

Note that thickness restriction applies only to the input space $\mathcal{Y}$, while the output space $\mathcal{X}$ is left entirely unrestricted. For example the committee spaces $\mathcal{X}_L^{\mathcal{K}}$ from Example 3 is not thick; since it is defined by an affine feasibility restriction, it has dimension $|\mathcal{K}| - 1$, not $|\mathcal{K}|$. Nevertheless, Theorem 1 still applies to Multi-Winner Approval Voting under Constraints, because the input space is $\mathcal{Y} = \{\pm 1\}^{\mathcal{K}}$. In the traditional setting in which $\mathcal{X} = \mathcal{Y}$, the thickness assumption obviously applies to the output space as well. In this case, Theorem 1 simplifies as follows.

Corollary 1 *Let $\mathcal{X}$ be a thick judgement space. An aggregation rule $F : \Delta(\mathcal{X}) \rightrightarrows \mathcal{X}$ satisfies* ESME*,* Continuity*, and* Reinforcement *if and only if F is the median rule.*

By Proposition 2, the statements of Theorem 1 and Corollary 1 remain true if Reinforcement is replaced by Judgement Consistency. All three axioms are necessary for the characterization. For example:

- If $\phi : [-1, 1] \longrightarrow \mathbb{R}$ is increasing and continuous, then the additive majority rule F_ϕ satisfies ESME and Continuity. But F_ϕ does not satisfy Reinforcement unless it is the median rule.

[9] Proposition 2 suggests that Judgement Consistency is logically stronger than Reinforcement. But Reinforcement depends on a *variable* population of voters, whereas Judgement Consistency can still be applied when the population of of voters is *fixed*. Since our framework assumes a variable population from the beginning, it somewhat obscures this distinction.

- All scoring rules satisfy Reinforcement and Continuity (Myerson, 1995), but typically violate ESME.

- Let $>$ be an arbitrary strict order on $\mathcal{X}$. One can construct a single-valued refinement of the median rule which satisfies ESME and Reinforcement and breaks any tie by choosing the $>$-maximal element. But this rule does not satisfy Continuity.

In the special case in which $\mathcal{X} = \mathcal{Y}$ is the space of linear orderings, Theorem 1 yields a counterpart to the seminal contribution by Young and Levenglick (1978). As we already mentioned in the introduction Young and Levenglick characterize the median rule for such spaces by three assumptions: Condorcet Consistency, Neutrality and Reinforcement. While their Reinforcement axiom is exactly the same as ours, the other axioms are not quite comparable. Condorcet Consistency is somewhat weaker than SME (hence *a fortiori* weaker than ESME) but not entirely, since it also deals with majority ties.[10] Neutrality is the standard axiom of a symmetry in alternatives. Since any additive majority rule is neutral, in view of Proposition 1, Neutrality is implied here by ESME plus Continuity. It is not quite implied by ESME alone, since non-neutral selections from AMRs would satisfy ESME as well. Conceptually, a lot of Neutrality is built into ESME via its symmetric treatment of issues.

The three axioms in the Young-Levenglick theorem are meaningful for general judgment aggregation contexts, with Neutrality understood as invariance to any symmetries of the context (input and output spaces) under permutations of issues. However, only rarely will they suffice to *uniquely characterize* the median rule, simply because in most contexts, there will be few if any symmetries to exploit.

Mathematically, the two results and their proofs are very different. Young and Levenglick's proof is a tour de force that strongly exploits the special combinatorial features of the permutation polytope $\text{conv}(\mathcal{X}_{\mathcal{A}}^{\text{rk}})$. By contrast, our proof of Theorem 1 needs to effectively sidestep the combinatorial structure of the context. Even equipped with Proposition 1, this requires significant work because the intended generality precludes the use of arguments that exploit special properties of the combinatorial structure of a particular context. A proof sketch of Theorem 1 will be provided at the end of section 5.

There is considerable discussion on various versions of the median rule in the more mathematically oriented literature; see, for example, Chapter 5 of the monograph of Day and McMorris (2003), and also Monjardet (2008). Axiomatizations appear largely confined to *median spaces* defined in Nehring and Puppe (2007) as an adaption of "median graphs" to judgment aggregation spaces. In particular, McMorris et al. (2000) provide a characterization of the median rule in median graphs/spaces based on a local Condorcet condition and Reinforcement.[11]

[10] Say that a view $\mathbf{x}$ is *Condorcet dominant* if, for all $k \in K$, $\widetilde{\mu}_k \cdot x_k \geq 0$. A rule F is *Condorcet consistent* if F is equal to the set of all Condorcet dominant views whenever that set is non-empty.

[11] However, in median spaces, the median rule is characterized by Condorcet consistency alone. Indeed, as shown in Nehring and Puppe (2007), the median spaces are exactly the spaces in which this is the case.

5 Extension to weighted judgement contexts

The formulation of judgement aggregation in Section 2 implicitly gave the same weight to the voters' opinions on all issues. But sometimes such "equal weighting" is not appropriate. For example, in a truth-functional aggregation problem, we may wish to give a higher weight to the voters' opinions about the premises than their opinions about the conclusions. The most extreme form of this is the "premise-based" aggregation rule, which aggregates the voters' views on each premise by majority vote, but completely ignores their opinions about the conclusions; instead, the collective opinion about each conclusion is logically derived from the majoritarian opinons on the premises. At the opposite extreme is the "conclusion-based" aggregation rule, which aggregates the voters' opinions on each conclusion by majority vote, and mostly ignores their opinions about the premises, except when these opinions can be aggregated in a manner which is logically consistent with the majority opinions about the conclusions. Between these extremes, there are rules which give greater or lesser weight to the voters' views on different premises and conclusions.

There are other judgement aggregation problems where one might want to assign different weight to the voters' opinions on different issues. Indeed, only if a problem had a high amount of "symmetry" (e.g. aggregation of rankings) would there be a strong *a priori* reason to assign the same weight to all issues. For this reason, we now introduce a ***weight vector*** $\boldsymbol{\lambda} = (\lambda_k)_{k\in\mathcal{K}}$, where $\lambda_k > 0$ is the "weight" which we assign to the voters' opinions on issue k. Roughly speaking, λ_k would be large if we were very unwilling to overrule the majority opinion in issue k. Conversely, λ_k would be small if we were quite ready to overrule this opinion, if this was necessary to achieve a coherent collective view. A ***weighted judgement context*** is a quadruple $\boldsymbol{\mathcal{C}} := (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$, where $\mathcal{K}$ is a (finite) set of issues, $\boldsymbol{\lambda} \in \mathbb{R}_+^{\mathcal{K}}$ is a weight vector, and $\mathcal{X} \subseteq \mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}}$, with $\mathcal{X}$ being the set of "admissible collective views", and $\mathcal{Y}$ being the set of "admissible individual views". (In particular, an *unweighted* judgement context of the kind considered in Sections 2 to 4 can be represented by setting $\boldsymbol{\lambda} = (1, 1, \ldots, 1)$.) A ***judgement problem*** is an ordered pair $(\boldsymbol{\mathcal{C}}, \mu)$, where $\boldsymbol{\mathcal{C}}$ is a (weighted) judgement context, and $\mu \in \Delta(\mathcal{Y})$ is a profile. For any $\mathbf{x} \in \mathcal{X}$ and $q \in [0, 1]$, we now define

$$\gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}}(q) \quad := \quad \sum \{\lambda_k \,;\; k \in \mathcal{K} \text{ and } x_k\, \widetilde{\mu}_k \geq q\}. \tag{6}$$

This yields a non-increasing function $\gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}} : [0, 1] \longrightarrow \mathbb{R}$. If $\boldsymbol{\mathcal{C}}$ is an *unweighted* judgement context (i.e. $\boldsymbol{\lambda} = (1, \ldots, 1)$), then formula (6) reduces to the definition of $\gamma_{\mathbf{x}}^{\mu}$ from Section 3. We say $\mathbf{x} \in \mathcal{X}$ is ***supermajority efficient*** (SME) for the judgement problem $(\boldsymbol{\mathcal{C}}, \mu)$ if there does not exist any $\mathbf{z} \in \mathcal{X}$ such that $\gamma_{\mu,\mathbf{z}}^{\boldsymbol{\lambda}}(q) \geq \gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}}(q)$ for all $q \in [0, 1]$, with strict inequality for some q. A judgement aggregation rule $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is ***supermajority efficient*** on $\boldsymbol{\mathcal{C}}$ if, for any $\mu \in \Delta(\mathcal{Y})$, every element of $F(\mu)$ is supermajority efficient for $(\boldsymbol{\mathcal{C}}, \mu)$.

If $\phi : [-1, 1] \longrightarrow {}^*\mathbb{R}$ is a gain function, then the ***additive majority rule*** on $\boldsymbol{\mathcal{C}}$ is the correspondence $F_\phi : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ defined as follows:

$$\text{for all } \mu \in \Delta(\mathcal{Y}), \qquad F_\phi(\mu) \quad := \quad \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmax}} \left(\sum_{k\in\mathcal{K}} \lambda_k\, \phi(x_k\, \widetilde{\mu}_k) \right). \tag{7}$$

In particular, the *median rule* on $\mathcal{C}$ is defined by

$$\begin{aligned} \text{Median}\,(\mathcal{C},\mu) &:= \operatorname*{argmax}_{\mathbf{x}\in\mathcal{X}} (\mathbf{x} \underset{\lambda}{\bullet} \widetilde{\boldsymbol{\mu}}), \qquad \text{for all } \mu \in \Delta(\mathcal{Y}), \\ \text{where}\quad \mathbf{x} \underset{\lambda}{\bullet} \widetilde{\boldsymbol{\mu}} &:= \sum_{k\in\mathcal{K}} \lambda_k\, x_k\, \widetilde{\mu}_k, \qquad \text{for all } \mathbf{x}\in\mathcal{X}. \end{aligned} \tag{8}$$

If $\mathcal{C}$ is an unweighted judgement context (i.e. $\boldsymbol{\lambda} = (1,\ldots,1)$), then this reduces to the "unweighted" median rule defined by formula (2). For any $\mathbf{x},\mathbf{y} \in \{\pm 1\}^{\mathcal{K}}$, we define their *$\boldsymbol{\lambda}$-weighted distance* by $d_{\boldsymbol{\lambda}}(\mathbf{x},\mathbf{y}) := \sum\{\lambda_k;\quad k \in \mathcal{K} \text{ and } x_k \neq y_k\}$. The median rule (8) can be equivalently defined as selecting the view(s) in $\mathcal{X}$ minimizing the average $\boldsymbol{\lambda}$-weighted distance to the views of the voters. In Section 6, we will provide an axiomatic characterization of the weighted median rule (8), similar to Theorem 1. First, we will motivate the theory of weighted judgement contexts in general —and the weighted median rule in particular —with two interesting applications: *assignment problems* and *uniform decisions.* But an impatient reader can skip directly to Section 6 without loss of logical continuity.

5.1 Assignment problems

Consider a group of voters who need to assign different candidates to different of positions such as the positions in a cabinet. There is a set $\mathcal{A} := \{1,\ldots,A\}$ of "candidates" and a set $\mathcal{B} := \{1,\ldots,B\}$ of "positions", with $A \geq B$.[12] (As in the matching literature, there are many different possible interpretation; for example, "positions" could be unique resources such as organ transplants, and "candidates" could be possible recipients. Judgment aggregation might be required when different group members entertain different standards of fair allocation).

Assignments can be described in terms of an issue space $\mathcal{K} = \mathcal{A} \times \mathcal{B}$, with the issue (a,b) addressing the question: "should candidate a hold position b"? Feasibility requires that any position be filled by exactly one candidate, and that any candidate can fill at most one position. Feasible assignments can thus be described as $B-$tuples $(\!(a_1,\ldots,a_B)\!)$ saying that candidate a_b is assigned to position b. More explicitly in issue space, the tuple $(\!(a_1,\ldots,a_B)\!)$ refers to the view $\mathbf{x} \in \{\pm 1\}^{\mathcal{A}\times\mathcal{B}}$ such that $x_{ab} = 1$ iff $a = a_b$, and $x_{ab} = -1$ iff $a \neq a_b$. The feasible output space is the set of all such judgments $\mathcal{X}^{\text{asgn}}_{\mathcal{A}\times\mathcal{B}}$. By contrast, we will allow input judgments to be unrestricted approval judgments for each position; thus $\mathcal{Y} = \{\pm 1\}^{\mathcal{A}\times\mathcal{B}}$.

Generally, different positions will differ in their "importance", so it will be natural to assign weights $\lambda_{a,b}$ of the form $\lambda_{a,b} = \overline{\lambda}_b$, where $\overline{\lambda}_b$ reflects the importance of position b. This defines a judgement context $\mathcal{C} = \left(\mathcal{A}\times\mathcal{B}, \boldsymbol{\lambda}, \mathcal{X}^{\text{asgn}}_{\mathcal{A}\times\mathcal{B}}, \{\pm 1\}^{\mathcal{A}\times\mathcal{B}}\right)$. In such a context, the median rule selects the assignment $\mathbf{x} = (\!(a_1,\ldots,a_B)\!)$ that maximizes the weighted sum

$$\sum_{b\in\mathcal{B}} \overline{\lambda}_b\, \widetilde{\mu}_{(a_b,b)}.$$

[12] This problem has been considered in particular by Emerson (2016). The median rule can be viewed as an Approval Voting counterpart of sorts to Emerson's "matrix vote". See also http://www.deborda.org/faq/voting-systems/what-is-the-matrix-vote.html

To illustrate the role of the weights, consider, as a simple example, a profile of input views μ all of which assess candidate quality as independent of position, i.e. $\widetilde{\mu}_{a,b} = \widetilde{\mu}_{a,b'}$ for all $a \in \mathcal{A}$, and $b, b' \in \mathcal{B}$. In such profiles, supermajority efficient rules such as the weighted median rule (8) will assign the better candidates to the more important positions. Likewise, in arbitrary profiles, if the weight for some position b^* is much larger than that for any other position, the median rule fills that position by the candidate a with the highest majority support $\widetilde{\mu}_{(a,b^*)}$ (up to, possibly, near ties).

5.2 Uniform Decisions

The following "uniform decision" model is in fact a scheme of examples generating more complex judgment aggregation contexts from simpler ones. We first present the formal scheme, and then illustrate three types of applications referred to as *Heterogeneous Cases*, *Missing Information*, and *Multiple Criteria*.

Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a "base" judgement context. Let $\mathcal{S}$ be an abstract set of "instances"; in the three applications, the elements of $\mathcal{S}$ will be interpreted as "cases", "states" or "criteria". Let $\overline{\boldsymbol{\lambda}} := (\overline{\lambda}_s)_{s\in\mathcal{S}}$ be a vector assigning a "weight" to each instance; in the three applications, these weights will be interpreted as the frequencies of the cases, the probabilities of the states, or the relative importance of the criteria. (We will provide more detail below.) While input judgments are made instance-wise, these form the basis for a single output judgment that governs all instances uniformly. This situation can be described formally by a "uniform" judgement context $\widehat{\boldsymbol{\mathcal{C}}} = (\widehat{\mathcal{K}}, \widehat{\boldsymbol{\lambda}}, \widehat{\mathcal{X}}, \widehat{\mathcal{Y}})$ defined as follows.

(i) $\widehat{\mathcal{K}} := \mathcal{K} \times \mathcal{S}$;

(ii) $\widehat{\mathcal{X}} := \{(\mathbf{x}, \ldots, \mathbf{x})\,;\, \mathbf{x} \in \mathcal{X}\}$;

(iii) $\widehat{\mathcal{Y}} := \mathcal{Y} \times \cdots \times \mathcal{Y}$;

(iv) For all $k \in \mathcal{K}$ and $s \in \mathcal{S}$, $\widehat{\lambda}_{k,s} := \overline{\lambda}_s\, \lambda_k$.

Part (iii) allows input judgments to vary independently across instances, while (ii) says that output judgments must be constant across instances. Part (iv) says that the weight of a composite issue (k, s) rescales the base weight λ_k by the instance weight $\overline{\lambda}_s$.

5.2.1 Heterogeneous Cases

To flesh out the first interpretation in terms of heterogeneous instances, consider the very simplest judgment aggregation problem given by a single yes-no issue, i.e. $|\mathcal{K}| = 1$.

For example, a group needs to decide on whether to establish a rule which permits or forbids some types of action or behaviour. Concrete examples include traffic laws, workplace codes of conduct, or safety regulations at a public swimming pool. More contentious examples include restrictions on libel, hate speech, obscenity, or incitement.[13] For this rule

[13] See Miller (2013) for a recent contribution to this theme in the judgment aggregation literature.

to be simple, unambiguous, and enforceable in a non-arbitrary and independently verifiable way, it must be based on a relatively coarse description of the action. Of course, in each particular case, the "correct" or "just" decision may depend on some finer details of the action (and surrounding context). But even if these details are observable in principle, it is not feasible to explicitly condition the actual decision on them, e.g. for reasons of complexity or non-verifiability.

So the best the group can do is to get it right "on balance" across comparable cases $s \in \mathcal{S}$. Voters give their judgments about the right decision in each case. The uniform group decision is then to be taken on the basis of the entire vector of majorities $(\widetilde{\mu}_s)_{s\in\mathcal{S}}$. The cases could be actual or hypothetical. For actual cases, the weights $\overline{\lambda}_s$ would naturally reflect their frequency of occurence. For hypothetical cases, the weights $\overline{\lambda}_s$ would naturally reflect their 'relevance' or 'representativeness'. Additive majority rules yields a positive uniform decision just in case

$$\sum_{s\in\mathcal{S}} \overline{\lambda}_s \, \phi\left(\widetilde{\mu}_s\right) \geq 0.$$

Hence, the median rule in particular yields a positive uniform decision iff

$$\sum_{s\in\mathcal{S}} \overline{\lambda}_s \, \widetilde{\mu}_s \geq 0,$$

i.e. just in case the weighted average of majorities across comparable cases is non-negative.

5.2.2 Missing Information

The uniform decision model can also be applied to situations in which the group is missing information relevant to the judgment task. For example, in foreign policy, macroeconomic management, and environmental regulation, the right course of action may depend on information which is unavailable at the moment the decision must be made. To pick a prominent recent example, during its 2014 independence referendum, Scotland confronted uncertainty about future petroleum prices and its ability to join the E.U. as an independent state. It is quite legitimate —and may lead to better decisions —to make this missing information explicit in the collective decision procedure by representing it as an unknown "state of nature" $s \in \mathcal{S}$, so that each voter submits her judgment contingent on the state. In a binary (single-issue) decision problem, analogously to Example 5.2.1 above, the median rule would base the decision on the *expected majority margin* $\sum_{s\in\mathcal{S}} p_s\widetilde{\mu}_s$.

The relevant probability weights $\overline{\lambda}_s = p_s$ could be obtained in various ways. For example, they might be obtained by some judgment aggregation rule from the voters themselves. Or, the group may delegate this judgment to an 'outside authority', for example to betting markets.[14]

This can be extended beyond a single issue, for instance to the ranking of more than two possible courses of action. The base context would then be given by the space of

[14] For a provocative and ambitious advocacy of using betting markets to "vote on values, but bet on beliefs", see Hanson (2013).

rankings $\mathcal{X}_{\mathcal{A}}^{\text{rk}}$ from Example 1. The median rule would select the ranking(s) $\succ$ with the highest support *ex ante*, as measured by the sum

$$\sum_{\substack{a,b\in\mathcal{A}\\ a\succ b}} \widehat{\mu}_{ab}, \tag{9}$$

where, for all $a, b \in \mathcal{A}$, we defined

$$\widehat{\mu}_{ab} \quad := \quad \sum_{s\in\mathcal{S}} p_s\, \widetilde{\mu}_{ab,s}. \tag{10}$$

The median rule thus yields an *extension of the Kemeny rule to uncertainty.* Note that the *ex ante* comparison depends as much on the size of the majorities as on their sign.

5.2.3 Multiple Criteria.

Applied to elections for public office, the standard ranking model can be interpreted as trying to determine the (impartially best) candidate based on an overall comparative evaluation of candidates in terms of their "suitability for office".

It might be argued that this gives too much room for subjective impressions of personal appeal of candidates, and that "suitability for office" can be more soundly elicited by explicitly taking account of its multi-criterial nature, on the view that "suitability for office" consists of a combination of rather distinct, identifiable qualities such as leadership, integrity, judgment, etc.. The multi-criterion conception of candidate merit can be captured by the uniform decision model as follows. Let the base context again be the standard ranking context $\mathcal{X}_{\mathcal{A}}^{\text{rk}}$ from Example 1, and let each s represent a different criterion. An input judgment $\mathbf{y}$ consists of an $\mathcal{S}$-tuple of rankings $(\mathbf{y}_s)_{s\in\mathcal{S}}$, with $\mathbf{y}_s$ representing the ranking of candidates in terms of criterion s. The uniform output judgment $\mathbf{x}$ represents the overall group ranking to be determined. It is based on weight vector $\overline{\boldsymbol{\lambda}} \in \mathbb{R}_+^{\mathcal{S}}$ describing the relative importance of these criteria. These can be deteremined in different ways: they could be determined concurrently by the group itself, by a separate committee or at an earlier 'constitutional' stage at which the general requirements for the office where determined.[15]

In this setting, the median rule selects the ranking with highest overall majority support (9), where again the overall majority support $\widehat{\mu}_{ab}$ for ranking of a over b is given the weighted average (10) . In this manner, the median rule thus yields an *extension of the Kemeny rule to multiple criteria.*

The general approach to multi-criterion evaluation just outlined here is not premised on the particular, comparative format evaluation in terms of rankings; an alternative format of interest is non-comparative in terms of "grades". In this vein, Balinski and Laraki (2010, ch. 21) introduce a multi-criteria majority grading rule[16] which – setting aside the

[15]Switching the setting, a hiring committee or university department may be tasked to evaluate an applicant for an open faculty position in terms of research, teaching and service, and the weights of these might be predetermined by standing university policy.

[16]Their term is "multi-criteria majority judgment".

treatment of tied grades – is equivalent to the weighted median rule in the uniform decision model for a base context that describes grading as follows.

Let $(G, >)$ be a finite, linearly ordered set, each element being interpreted as a grade. Grading can be defined as a judgment context $(G, \mathcal{X}_G, \mathcal{X}_G, \boldsymbol{\lambda})$, where the issues $g \in G$ are interpreted as whether or not the object of evaluation achieves at least grade g, and the set of admissible grade assignments is given by $\mathcal{X}_G = \{\mathbf{x} \in \{\pm 1\}^G;\ x_g \geq x_{g'}$ whenever $g \leq g'\}$, with typical element $(1, \ldots 1, 0, \ldots, 0)$. The input space and output space agree, $\mathcal{Y}_G = \mathcal{X}_G$; the issue weights $\boldsymbol{\lambda}$ turn out to be immaterial.

6 Characterization of the weighted median rule

To obtain a weighted generalization of Theorem 1, we must consider weighted combinations of weighted contexts. These are given by weighted contexts of the form $\widehat{\mathcal{C}} = (N \cdot \mathcal{K}, \widehat{\boldsymbol{\lambda}}, \mathcal{X}^N, \mathcal{Y}^N)$, where $\widehat{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\lambda}}_1, \ldots, \widehat{\boldsymbol{\lambda}}_N) \in \mathbb{R}^{N \cdot \mathcal{K}}$ is *proportional* to $\boldsymbol{\lambda}$; that is, for for each $n \in [1, .., N]$, $\widehat{\boldsymbol{\lambda}}_n = c_n \boldsymbol{\lambda}$, for some $c_n > 0$. So, for a proportional weight vector $\widehat{\boldsymbol{\lambda}}$, the relative weights *within* the set of basic issues $\mathcal{K}$ are the same (original ones) in each instance, while different instances may be assigned different relative weights reflected in the scaling factor c_n. The potential differences in the scaling factor may have different origins, as in the uniform decision model of Section 5.2. In particular, they could reflect differences in frequency, probability or relative importance. We are now in a position to state the axiom of Weighted Ensemble Supermajority Efficiency.

WESME. For any set of instances N, any vector of weights $\widehat{\boldsymbol{\lambda}}$ proportional to $\boldsymbol{\lambda}$, and any profile $\mu \in \Delta\left(\mathcal{Y}^N\right)$, any element of $F\left(\mu^1\right) \times \cdots \times F\left(\mu^N\right)$ is SME for $(\widehat{\mathcal{C}}, \mu)$.

Here is the extension of Proposition 1 to weighted contexts.

Proposition 3 *Let F be a judgement aggregation rule on a weighted judgement context $\mathcal{C}$. If F satisfies* WESME *and* Continuity*, then F is an additive majority rule like* (7).

Note that the consideration of proportional weights is needed only to deal with the case of irrational-valued weight vectors $\boldsymbol{\lambda}$ in the base context. If all weights are rational, then it is sufficient to confine attention to weight vectors $\widehat{\boldsymbol{\lambda}}$ such that $\widehat{\boldsymbol{\lambda}}_n = \boldsymbol{\lambda}$ for all $n \in [1 \ldots N]$. (See Nehring and Pivato (2018) for more information.)

To obtain an axiomatic characterization of the median rule on a weighted judgement context $\mathcal{C}$, we will need $\mathcal{C}$ to satisfy some structural conditions. For any $\mu \in \Delta(\{\pm 1\}^{\mathcal{K}})$, let $\mathbf{x}^\mu$ be the *majority ideal*, as defined below formula (1) above. A judgement space $\mathcal{Y}$ is *McGarvey* if, for all $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$, there is some $\mu \in \Delta(\mathcal{Y})$ such that $\mathbf{x}^\mu = \mathbf{x}$.[17] For example, the spaces $\mathcal{X}_{\mathcal{A}}^{\mathrm{rk}}$ and $\mathcal{X}_{\mathcal{A}}^{\mathrm{eq}}$ (Examples 1 and 2) are McGarvey, as are many other commonly occuring judgement spaces; see Nehring and Pivato (2011) for many more examples. Clearly, any McGarvey space is thick, and any superset of a McGarvey space is also McGarvey.

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \{\pm 1\}^{\mathcal{K}}$. We say that $\mathbf{y}$ is *between* $\mathbf{x}$ and $\mathbf{z}$ if, for any $k \in \mathcal{K}$ such that $x_k = z_k$, we also have $y_k = z_k$ (and hence, $y_k = x_k$). Now let $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$ be a judgement

[17]Equivalently, $\mathcal{Y}$ is McGarvey if and only if the zero vector $\mathbf{0}$ lies in the topological interior of $\mathrm{conv}(\mathcal{Y})$.

space. Say that $\mathbf{x}, \mathbf{z} \in \{\pm 1\}^{\mathcal{K}}$ are ***near*** if there is no $\mathbf{y} \in \mathcal{X} \backslash \{\mathbf{x}, \mathbf{z}\}$ such that $\mathbf{y}$ is between $\mathbf{x}$ and $\mathbf{z}$. We will say that $\mathcal{X}$ is ***distal*** if there exist some $\mathbf{x}, \mathbf{z} \in \mathcal{X}$ such that $\mathbf{x}$ is near $\mathbf{z}$ and $d(\mathbf{x}, \mathbf{z}) \geq 3$. Heuristically, this means that the elements of $\mathcal{X}$ are not packed tightly together everywhere. For any $\mathbf{z} \in \{\pm 1\}^{\mathcal{K}} \backslash \mathcal{X}$, let $\mathcal{X}_{\mathbf{z}}$ be the set of views in $\mathcal{X}$ that are near to $\mathbf{z}$; heuristically, these are the "best admissible approximations" to $\mathbf{z}$. We say that $\mathcal{X}$ is ***rugged*** if there exists $\mathbf{z} \in \{\pm 1\}^{\mathcal{K}} \backslash \mathcal{X}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{X}_{\mathbf{z}}$ such that $d(\mathbf{x}, \mathbf{z}) \neq d(\mathbf{y}, \mathbf{z})$.

Both distality and ruggedness are conditions on the combinatorial geometry of $\mathcal{X}$, and typically easy to verify. Ruggedness seems to be satisfied in the great majority of cases, while distality is somewhat more restrictive. Indeed, distality "almost" implies ruggedness. That is, suppose that there exist some $\mathbf{x}, \mathbf{z} \in \mathcal{X}$ such that $\mathbf{x}$ is near $\mathbf{z}$ and $d(\mathbf{x}, \mathbf{z}) \geq 3$, and assume additionally that some such $d(\mathbf{x}, \mathbf{z})$ is odd. Consider any $\mathbf{y} \in \{\pm 1\}^{\mathcal{K}} \backslash \{\mathbf{x}, \mathbf{z}\}$ such that $\mathbf{y}$ is between $\mathbf{x}$ and $\mathbf{z}$. Since $\mathbf{x}$ is near $\mathbf{z}$, $\mathbf{y} \in \{\pm 1\}^{\mathcal{K}} \backslash \mathcal{X}$. Since $d(\mathbf{x}, \mathbf{z}) = d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$ and $d(\mathbf{x}, \mathbf{z})$ is odd, $d(\mathbf{x}, \mathbf{y}) \neq d(\mathbf{y}, \mathbf{z})$, verifying the ruggedness of $\mathcal{X}$. We now come to the second main result of this paper.

Theorem 2 *Let* $(\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ *be a weighted judgement context, and suppose that either* (a) $\mathcal{X}$ *is rugged and* $\mathcal{Y}$ *is McGarvey, or* (b) $\mathcal{X}$ *is distal and* $\mathcal{Y}$ *is thick. Let* $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ *be a judgement aggregation rule. Then* F *satisfies* WESME*,* Continuity*, and* Reinforcement *if and only if* F *is the median rule* (8)*.*

Thus (weighted) median rules (8) are both (weighted) additive majority rules and scoring rules in the sense of Myerson (1995) or Dietrich (2014). Both types of judgment aggregation rules are arguably fundamental, but they are characterized by distinct separability conditions: separability in issues versus separability in individuals. (Weighted) median rules are separable in both of these senses. Indeed, since scoring rules satisfy Reinforcement in general, Theorem 2 provides general conditions under which weighted median rules are the *only* rules with this "double separability".

The presuppositions of Theorem 2 are satisfied in many applications of interest, but they are more restrictive than those of Theorem 1. We illustrate their broad applicability in a variety of examples before providing a nonexample.

1. As already mentioned, the spaces of rankings (Example 1) and of classifiers (Example 2) are McGarvey. They are also easily seen to be rugged. Furthermore, the latter is distal, while the former is not.

2. In the assignment problems of Section 5.1, the input space is trivially McGarvey as it is an approval space. For $B \geq 3$, it is rugged. To see this, consider for the assignments $\mathbf{z} := \langle\!\langle a_1, a_1, a_2 \rangle\!\rangle$, $\mathbf{x} := \langle\!\langle a_1, a_3, a_2 \rangle\!\rangle$ and $\mathbf{y} := \langle\!\langle a_1, a_2, a_3 \rangle\!\rangle$. Note that $\mathbf{x}$ and $\mathbf{y}$ are feasible, while $\mathbf{z}$ is not. Evidently, $\mathbf{x}$ and $\mathbf{y}$ are adjacent to $\mathbf{z}$; the triple of views verifies ruggedness since $d(\mathbf{z},\mathbf{x}) = 4$ while $d(\mathbf{z}, \mathbf{y}) = 8$.

3. In the uniform decision problems of Section 5.2, it follows from elementary linear algebra to see that $\widehat{\mathcal{Y}}$ is thick if and only if $\mathcal{Y}$ is thick. On the other hand, for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, $d((\mathbf{x}, \ldots \mathbf{x}), (\mathbf{y}, \ldots, \mathbf{y})) = |S|\, d(\mathbf{x}, \mathbf{y})$. Thus $\mathcal{X}$ is distal whenever $|S| \geq 3$.

It is not surprising that Theorem 2 needs the domain $\mathcal{Y}$ to satisfy a richness condition such as thickness or the McGarvey property. It is perhaps more surprising that the codomain $\mathcal{X}$ must satisfy a structural condition like ruggedness or distality. To see why these conditions are necessary, consider the following class of additive majority rules. For any $\alpha > 0$, we define $\phi^\alpha : [-1,1] \longrightarrow \mathbb{R}$ by $\phi_\alpha(r) := \mathrm{sign}(r) \cdot |r|^\alpha$. The corresponding additive majority rule F_{ϕ^α} is called a ***homogeneous*** rule. Note that the median rule is just the homogeneous rule with $\alpha = 1$. Now let $\mathcal{K}$ be a finite set, and define $\mathcal{X}^K_{L,M} := \{\mathbf{x} \in \{\pm 1\}^{\mathcal{K}};\ L \leq \#\mathbf{x} \leq M\}$. If $L < M$, this space is thick, and if $L < \frac{K}{2} < M$, then it is McGarvey.[18] But Theorem 2 fails on this space. Its normative assumptions fail to single out the median rule uniquely, but are satisfied by any homogeneous rule.

Proposition 4 *Let $K \geq M \geq L \geq 0$, let $\boldsymbol{\lambda} \in \mathbb{R}^{\mathcal{K}}_+$ be a weight vector, and let $\boldsymbol{\mathcal{C}} := (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}^K_{L,M}, \mathcal{X}^K_{L,M})$. For any $\alpha > 0$, the rule F^α satisfies* Reinforcement*,* Continuity*, and* WESME *on $\boldsymbol{\mathcal{C}}$. If $\boldsymbol{\lambda}$ is the uniform weight vector, then F^α is equivalent to the median rule on $\boldsymbol{\mathcal{C}}$. But if $\boldsymbol{\lambda}$ not uniform and $\alpha \neq 1$, then F^α is not the median rule on $\boldsymbol{\mathcal{C}}$.*

The structural assumptions of Theorem 2 fail for $\mathcal{X}^K_{L,M}$. To see that $\mathcal{X} = \mathcal{X}^{\mathcal{K}}_{0,M}$ is not rugged, for example, one simply notes that, for any $\mathbf{z} \notin \mathcal{X}$, $\mathbf{x}$ is near $\mathbf{z}$ iff $\#\mathbf{x} = M$ and if, for all $k \in \mathcal{K}$, $x_k = 1$ implies $z_k = 1$. Thus, for any $\mathbf{x}$ near $\mathbf{z}$, we have $d(\mathbf{x}, \mathbf{z}) = \#\mathbf{z} - M$, contradicting ruggedness. The failure of distality is verified easily as well.

Note that one can translate any judgment aggregation context with rational weights into an 'equivalent' judgment aggregation context with uniform weights by "cloning" issues in proportion to their weight. However, in this translation, due to the cloning, thickness of the input space tends to get lost. So one can use Proposition 4 to produce a non-thick counterexample to Theorem 1 in the unweighted case.

A sketch of the proofs. Theorems 1 and 2 are both results on judgment contexts and consequences of a more basic result on judgment aggregation rules, Theorem B.1, which appears in Appendix B. From Propositions 1 and 3, we know that any rule satisfying Continuity and (W)ESME is an additive majority rule F_ϕ, for some gain function ϕ. Thus, the key task in the proof of Theorem B.1 is to show that the identity function (or any linear function) on $[-1,1]$ is among these gain functions.

A possible strategy would be to try to show that any representing gain function was linear. If ϕ was real-valued, it would then be enough to show that ϕ satisfies the Cauchy functional equation, and whence deduce linearity. But this straightforward-looking strategy fails for two reasons. First of all, Propositions 1 and 3 do not guarantee that ϕ is real-valued, so one cannot appeal to the Cauchy functional equation. Second, the conclusions of Theorems 1 and 2 do not require that *all* representing gain functions ϕ are linear, only that *some* of them are. Indeed, Approval Voting on size-L committees provides

[18] $\mathcal{X}^K_{L,M}$ is often used to represent committee selection problems, as in Example 3: $\mathcal{K}$ is a set of potential "candidates", and the committee in question must have at least L and most M members. It also arises in certain resource allocation problems. But these interpretations are not relevant for Proposition 4.

a counterexample, in which the assumptions of Theorem 1 are satisfied, but any AMR agrees with the median rule, whatever the gain functions.

Hence this strategy is a nonstarter. Instead, the proof of Theorem B.1 shows in a couple of steps that any representing ϕ has the "linearity-like" property that $\phi(r+\epsilon)-\phi(r) = \phi(s+\epsilon)-\phi(r)$ for any r and s, and any sufficiently small ϵ; in effect, this says that ϕ has a "constant slope" property. This is enough to show that F_ϕ is the median rule, because it implies that a gain (or loss) of ϵ in the majority support on one issue can be exactly offset be a gain (or loss) of ϵ in the support on another issue.[19]

To show in turn that ϕ satisfies this "constant slope" property, we must study "perfect tie" profiles $\mu \in \Delta(\mathcal{Y})$ such that $F(\mu) = \{\mathbf{x}, \mathbf{y}\}$ for some $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. If the set of such "perfect tie" profiles is not empty, then it is a relatively open subset of an affine hyperplane (this is a consequence of Continuity and Reinforcement); we can then demonstrate the "constant slope" property by looking at how the values of $\phi(\widetilde{\boldsymbol{\mu}})$ change as μ moves around in this hyperplane. However, we also need $\mathbf{x}$ and $\mathbf{y}$ to differ in at least three coordinates —in effect, this is because we need the freedom to manipulate the r, s, and ϵ variables independently while preserving the perfect tie. In this case, let's say that $\{\mathbf{x}, \mathbf{y}\}$ is a *suitable pair*.

Theorem B.1 just *assumes* that a suitable pair exists for the given judgment aggregation rule F . So to derive Theorems 1 and 2 from Theorem B.1, we must provide conditions on the judgment context which guarantee the existence of a suitable pair , for any judgment aggregation rule satisfying the normative assumptions. This is done in Theorem B.2 in the Appendix in terms of a constructive condition of 'frangibility' on contexts.

If $\mathcal{X}$ is rugged and $\mathcal{Y}$ is McGarvey, then frangibility of the context follows from Lemma B.1 in the Appendix; meanwhile if $\mathcal{X}$ is distal and $\mathcal{Y}$ is thick, then frangibility of the context follows from Lemma B.2. Finally, Theorem 2 follows by combining Theorem B.2 with Lemmas B.1 and B.2.

For *unweighted* judgement contexts, a second route is available. If there exists a suitable pair, then we can argue as above. Otherwise, that is if $d(\mathbf{x}, \mathbf{y}) \leq 2$ whenever $F(\mu) = \{\mathbf{x}, \mathbf{y}\}$, then we can use a straightforward geometric argument (along with the three axioms) to show that F is the median rule. This is why Theorem 1 does not require the auxiliary structural conditions of Theorem 2. On the other hand, it is typically not the case that $d(\mathbf{x}, \mathbf{y}) \leq 2$ whenever $F(\mu) = \{\mathbf{x}, \mathbf{y}\}$. Thus, even in the unweighted case, the proof of Theorem 1 requires the full force of Theorem B.1.

Appendix A: Proofs of minor results

This appendix contains the proofs of Propositions 1 to 4.

Proof of Propositions 1 and 3. Both statements follow from Corollary 5.2 of Nehring and Pivato (2018). Let $\langle \mathcal{C} \rangle := \{\mathcal{C}^N;\ \ N \in \mathbb{N}\}$ (in the unweighted case of Section 3), and let $\langle \mathcal{C} \rangle := \{\mathcal{C}^{\mathbf{p}};\ \ \mathbf{p} \in \Delta^N \text{ and } N \in \mathbb{N}\}$ (in the weighted case of Section 5). Let $F^* := \{F^N\}_{N=1}^{\infty}$. Then in the terminology of Nehring and Pivato (2018), F^* is a

[19] This assumes uniform weights. In the case of nonuniform weights, the relevant statement is that a gain (or loss) of ϵ/λ_j in the majority support on issue j can be exactly offset be a gain (or loss) of ϵ/λ_k in the majority support on issue k.

judgement aggregation rule on $\langle \mathcal{C} \rangle$, and it satisfies the axiom of *Combination* (in either the weighted or unweighted case) and *Scale invariance* (in the weighted case). The axioms ESME (in the unweighted case) or WESME (in the weighted case) say that F^* satisfies the axiom of *Supermajority efficiency* on $\langle \mathcal{C} \rangle$. Finally, Continuity says that the rule F is continuous on $\Delta(\mathcal{Y})$. Thus, Corollary 5.2 of Nehring and Pivato (2018) says that F is an additive majority rule. □

The proof of Proposition 2 requires some preliminaries. Let $(\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a judgement context, and let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be a judgement aggregation rule. For any $\mu, \nu \in \Delta(\mathcal{Y})$, define $\mu \oplus \nu := \frac{1}{2}\mu + \frac{1}{2}\nu$. Consider the following, weaker versions of Reinforcement and Judgement Consistency.

Even Reinforcement: For any $\mu_1, \mu_2 \in \Delta(\mathcal{Y})$, if $F(\mu_1) \cap F(\mu_2) \neq \emptyset$, then $F(\mu_1 \oplus \mu_2) = F(\mu_1) \cap F(\mu_2)$.

Even Judgement Consistency: For any $\mu, \mu', \nu, \nu' \in \Delta(\mathcal{Y})$, and any views $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if (a) $\mathbf{y} \in F(\mu \oplus \nu)$, and (b) $\mathbf{x} \in F(\mu' \oplus \nu)$, and (c) $\mathbf{y} \notin F(\mu' \oplus \nu)$, and (d) $\mathbf{x} \in F(\mu \oplus \nu')$, then (e) $\mathbf{y} \notin F(\mu' \oplus \nu')$.

Lemma A.1 *Let $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be any judgement context, and let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be any judgement aggregation rule. If F satisfies* Even Judgement Consistency*, then F satisfies* Even Reinforcement*.*

Proof: Let $\mu_1, \mu_2 \in \Delta(\mathcal{Y})$.

Claim 1: *Let $\mathbf{y} \in \mathcal{X}$. If $\mathbf{y} \in F(\mu_1)$ and $\mathbf{y} \in F(\mu_2)$, then $\mathbf{y} \in F(\mu_1 \oplus \mu_2)$.*

Proof: Set $\mu = \nu = \mu_1$ and $\mu' = \nu' = \mu_2$. Then for any $\mathbf{x} \in \mathcal{X} \setminus \{\mathbf{y}\}$, Even Judgement Consistency says: if (a) $\mathbf{y} \in F(\mu_1)$, and (bd) $\mathbf{x} \in F(\mu_1 \oplus \mu_2)$, and (c) $\mathbf{y} \notin F(\mu_2 \oplus \mu_1)$, then $\mathbf{y} \notin F(\mu_2)$.

Taking the contrapositive, if $\mathbf{y} \in F(\mu_2)$, then one of the hypotheses (a), (bd), or (c) must be false. In particular, if $\mathbf{y} \in F(\mu_2)$ and $\mathbf{y} \in F(\mu_1)$, then

$$\text{either} \quad \mathbf{x} \notin F(\mu_1 \oplus \mu_2) \quad \text{or} \quad \mathbf{y} \in F(\mu_1 \oplus \mu_2). \tag{A1}$$

This holds for all $\mathbf{x} \in \mathcal{X} \setminus \{\mathbf{y}\}$.

Now, by contradiction, suppose $\mathbf{y} \in F(\mu_2)$ and $\mathbf{y} \in F(\mu_1)$, but $\mathbf{y} \notin F(\mu_1 \oplus \mu_2)$. Applying (A1), we obtain $\mathbf{x} \notin F(\mu_1 \oplus \mu_2)$ for all $\mathbf{x} \in \mathcal{X} \setminus \{\mathbf{y}\}$, which means that $F(\mu_1 \oplus \mu_2) = \emptyset$. This is a contradiction. The claim follows. ◇ Claim 1

Claim 2: *Let $\mathbf{y} \in \mathcal{X}$. If $F(\mu_1) \cap F(\mu_2) \neq \emptyset$, and $\mathbf{y} \notin F(\mu_2)$, then $\mathbf{y} \notin F(\mu_1 \oplus \mu_2)$.*

Proof: Let $\mathbf{x} \in F(\mu_1) \cap F(\mu_2)$. If we set $\mu' = \nu = \mu_2$ and $\mu = \nu' = \mu_1$, then Even Judgement Consistency says: if (a) $\mathbf{y} \in F(\mu_1 \oplus \mu_2)$, and (b) $\mathbf{x} \in F(\mu_2)$, and (c) $\mathbf{y} \notin F(\mu_2)$, and (d) $\mathbf{x} \in F(\mu_1)$, then (e) $\mathbf{y} \notin F(\mu_2 \oplus \mu_1)$.

Thus, $\mathbf{y} \notin F(\mu_2 \oplus \mu_1)$, as claimed. ◇ Claim 2

Combining Claims 1 and 2, we conclude that, if $F(\mu_1) \cap F(\mu_2) \neq \emptyset$, then $F(\mu_1 \oplus \mu_2) = F(\mu_1) \cap F(\mu_2)$. □

Lemma A.2 *Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be any judgement context, and suppose $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ satisfies* Continuity. *If F satisfies* Even Reinforcement*, then F satisfies* Reinforcement.

Proof: Let $\mu_1, \mu_2 \in \Delta(\mathcal{Y})$, and suppose $F(\mu_1) \cap F(\mu_2) \neq \emptyset$. Let $\mathbb{Q}_2$ be the set of dyadic rationals (that is, $\mathbb{Q}_2 := \{\frac{n}{2^k};\ n$ and k integers, $k > 0\}$).

Claim 1: $F(q\,\mu_1 + (1-q)\,\mu_2) = F(\mu_1) \cap F(\mu_2)$ *for all* $q \in \mathbb{Q}_2 \cap [0, 1]$.

Proof: Even Reinforcement implies that $F(\mu_1 \oplus \mu_2) = F(\mu_1) \cap F(\mu_2)$. It follows that $F(\mu_1 \oplus \mu_2) \cap F(\mu_1) = F(\mu_1) \cap F(\mu_2) \neq \emptyset$ and $F(\mu_1 \oplus \mu_2) \cap F(\mu_2) = F(\mu_1) \cap F(\mu_2) \neq \emptyset$. Thus, applying Even Reinforcement again, we deduce that $F[\mu_1 \oplus (\mu_1 \oplus \mu_2)] = F(\mu_1) \cap F(\mu_2)$ and $F[(\mu_1 \oplus \mu_2) \oplus \mu_2] = F(\mu_1) \cap F(\mu_2)$. But $\mu_1 \oplus (\mu_1 \oplus \mu_2) = \frac{3}{4}\mu_1 + \frac{1}{4}\mu_2$, while $(\mu_1 \oplus \mu_2) \oplus \mu_2 = \frac{1}{4}\mu_1 + \frac{3}{4}\mu_2$. Iterating this argument yields the claim, by induction. ◇ Claim 1

Now, $\mathbb{Q}_2$ is dense in $\mathbb{R}$. Thus, Continuity and Claim 1 imply that

$$F(r\,\mu_1 + (1-r)\,\mu_2) \supseteq F(\mu_1) \cap F(\mu_2), \qquad \text{for all } r \in [0, 1]. \tag{A2}$$

Claim 2: $F(r\,\mu_1 + (1-r)\,\mu_2) \subseteq F(\mu_1) \cap F(\mu_2)$ *for all* $r \in [0, 1]$.

Proof: Suppose $\mathbf{y} \notin F(\mu_1)$. We must show that $\mathbf{y} \notin F(r\,\mu_1 + (1-r)\,\mu_2)$ for all $r \in [0, 1]$.

By contradiction, let $\mathcal{R}_\mathbf{y} := \{r \in [0, 1];\ \mathbf{y} \in F(r\,\mu_1 + (1-r)\,\mu_2)\}$, and suppose $\mathcal{R}_\mathbf{y} \neq \emptyset$. By Continuity, 1 is not a cluster point of $\mathcal{R}_\mathbf{y}$ (because $\mathbf{y} \notin F(\mu_1)$). Thus, if $R := \sup(\mathcal{R})$, then $R < 1$. Now let $r \in \mathcal{R}_\mathbf{y}$. Find some $s \in (R, 1)$ such that $r/s = q \in \mathbb{Q}_2$.

Let $\nu := s\,\mu_1 + (1-s)\,\mu_2$. Then $\mathbf{y} \notin F(\nu)$, by definition of R. But $F(\nu) \supseteq F(\mu_1) \cap F(\mu_2)$, by statement (A2). Thus, $F(\nu) \cap F(\mu_2) \neq \emptyset$. Thus, Claim 1 implies that $F(q\,\nu + (1-q)\,\mu_2) = F(\nu) \cap F(\mu_2)$. But $q\,\nu + (1-q)\,\mu_2 = r\,\mu_1 + (1-r)\,\mu_2$ (because $r = q\,s$). Thus, we conclude that $F(r\,\mu_1 + (1-r)\,\mu_2) = F(\nu) \cap F(\mu_2)$. Hence $\mathbf{y} \notin F(r\,\mu_1 + (1-r)\,\mu_2)$, which contradicts the fact that $r \in \mathcal{R}_\mathbf{y}$.

To avert the contradiction, we must have $\mathcal{R}_\mathbf{y} = \emptyset$. This argument works for any $\mathbf{y} \notin F(\mu_1)$, and likewise any $\mathbf{y} \notin F(\mu_2)$. The claim follows. ◇ Claim 2

Statement (A2) and Claim 2 imply $F(r\,\mu_1 + (1-r)\,\mu_2) = F(\mu_1) \cap F(\mu_2)$ for all $r \in [0, 1]$. Thus, F satisfies Reinforcement. □

Proof of Proposition 2. If F satisfies Judgement Consistency, then it satisfies Even Judgement Consistency. Then Lemma A.1 says F satisfies Even Reinforcement. If F also satisfies Continuity then Lemma A.2 says it satisfies Reinforcement. □

Proof of Proposition 4. It is easy to verify that F^α satisfies WESME and Continuity. (Or see Proposition 3.4 and Corollary 5.2 from Nehring and Pivato (2018).) Thus, it remains to show that F^α satisfies Reinforcement.

For any weight vector $\boldsymbol{\lambda} = (\lambda_k)_{k\in\mathcal{K}}$, and any exponent $\alpha > 0$, we define the correspondence $F^\alpha_{\boldsymbol{\lambda}} : \Delta(\mathcal{X}^K_{L,M}) \rightrightarrows \mathcal{X}^K_{L,M}$ by setting

$$F^\alpha_{\boldsymbol{\lambda}}(\mu) \quad := \quad \operatorname*{argmax}_{\mathbf{x}\in\mathcal{X}} \left(\sum_{k\in\mathcal{K}} \lambda_k\, \phi_\alpha(x_k\, \widetilde{\mu}_k)\right), \quad \text{for all } \mu \in \Delta(\mathcal{X}^K_{L,M}). \tag{A3}$$

We also define $\boldsymbol{\lambda}^\alpha := (\lambda^\alpha_k)_{k\in\mathcal{K}}$.

Claim 1: *Let $\alpha, \beta > 0$, and let $\boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}^{\mathcal{K}}_+$ be weight vectors. If $\boldsymbol{\lambda}^\beta = \boldsymbol{\kappa}^\alpha$, then $F^\alpha_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M}, \mu) = F^\beta_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M}, \mu)$ for all $\mu \in \Delta(\mathcal{X}^K_{L,M})$.*

Proof: Let $\mu \in \Delta(\mathcal{X}^K_{L,M})$, and let $\mathcal{J} := \{k \in \mathcal{K};\ \widetilde{\mu}_k > 0\}$. There are now three cases.

Case 1. If $L \leq |\mathcal{J}| \leq M$, then $F^\alpha_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M}, \mu) = F^\beta_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M}, \mu) = \{\mathbf{x}\}$, where $\mathbf{x} \in \mathcal{X}^K_{L,M}$ is the view with $x_j = 1$ for all $j \in \mathcal{J}$ and $x_k = -1$ for all $k \in \mathcal{K} \setminus \mathcal{J}$.

Case 2. Suppose $|\mathcal{J}| > M$. Define $\boldsymbol{\eta} := \boldsymbol{\lambda}^{1/\alpha}$. Then we also have $\boldsymbol{\eta} = \boldsymbol{\kappa}^{1/\beta}$ (because $\boldsymbol{\eta}^\beta = \boldsymbol{\lambda}^{\beta/\alpha} = (\boldsymbol{\lambda}^\beta)^{1/\alpha} = (\boldsymbol{\kappa}^\alpha)^{1/\alpha} = \boldsymbol{\kappa}$, because $\boldsymbol{\lambda}^\beta = \boldsymbol{\kappa}^\alpha$ by hypothesis). By reordering the elements of $\mathcal{K} = [1 \dots K]$ if necessary, we can assume without loss of generality that

$$\eta_1\, \widetilde{\mu}_1 \ \geq\ \eta_2\, \widetilde{\mu}_2 \ \geq\ \cdots \ \geq\ \eta_J\, \widetilde{\mu}_J \ >\ 0 \ \geq\ \eta_{J+1}\, \widetilde{\mu}_{J+1} \ \geq\ \cdots \ \geq\ \eta_K\, \widetilde{\mu}_K. \tag{A4}$$

(Thus, $\mathcal{J} = [1 \dots J]$.) Now, for any $k \in \mathcal{K}$, observe that $\lambda_k\, \phi_\alpha(x_k\, \widetilde{\mu}_k) = \operatorname{sign}(x_k)\, \lambda_k\, |\widetilde{\mu}_k|^\alpha = \operatorname{sign}(x_k)\, |\eta_k\, \widetilde{\mu}_k|^\alpha$. Thus,

$$\sum_{k\in\mathcal{K}} \lambda_k\, \phi_\alpha(x_k\, \widetilde{\mu}_k) \quad = \quad \sum_{k\in\mathcal{K}} \operatorname{sign}(x_k)\, |\eta_k\, \widetilde{\mu}_k|^\alpha. \tag{A5}$$

Suppose $\eta_M\, \widetilde{\mu}_M > \eta_{M+1}\, \widetilde{\mu}_{M+1}$. Then the unique maximizer in $\mathcal{X}^K_{L,M}$ of the sum (A5) is the element $\mathbf{x} \in \mathcal{X}^K_{L,M}$ such that $x_m = 1$ for all $m \in [1 \dots M]$, while $x_k = -1$ for all $k \in [M+1 \dots K]$. Thus, definition (A3) yields $F^\alpha_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M}, \mu) = \{\mathbf{x}\}$. But by an identical argument, $\kappa_k\, \phi_\beta(x_k\, \widetilde{\mu}_k) = \operatorname{sign}(x_k)\, |\eta_k\, \widetilde{\mu}_k|^\beta$, so that

$$\sum_{k\in\mathcal{K}} \kappa_k\, \phi_\beta(x_k\, \widetilde{\mu}_k) \quad = \quad \sum_{k\in\mathcal{K}} \operatorname{sign}(x_k)\, |\eta_k\, \widetilde{\mu}_k|^\beta, \tag{A6}$$

so this sum is *also* uniquely maximized by $\mathbf{x}$, so $F^\beta_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M}, \mu) = \{\mathbf{x}\}$ also. Thus, $F^\alpha_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M}, \mu) = F^\beta_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M}, \mu)$, as claimed.

On the other hand, suppose $\eta_{N-1}\, \widetilde{\mu}_{N-1} > \eta_N\, \widetilde{\mu}_N = \eta_{N+1}\, \widetilde{\mu}_{N+1} = \cdots = \eta_P\, \widetilde{\mu}_P > \eta_{P+1}\, \widetilde{\mu}_{P+1}$ for some N, P with $N \leq M \leq P$. In this case, the sums (A5) and (A6) have more than one maximizer.[20] Even in this case, however, it is easy to see that they have exactly the *same* set of maximizers, so once again $F^\alpha_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M}, \mu) = F^\beta_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M}, \mu)$.

[20] To be precise, they have $\binom{P-N+1}{M-N+1}$ maximizers.

Case 3. Suppose $|\mathcal{J}| < L$. The argument is similar to *Case 2*. Again, assume without loss of generality that (A4) holds. If $\eta_L\,\widetilde{\mu}_L > \eta_{L+1}\,\widetilde{\mu}_{L+1}$, then by invoking equations (A5) and (A6), we see that $F^{\alpha}_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M},\mu) = F^{\beta}_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M},\mu) = \{\mathbf{x}\}$, where $\mathbf{x} \in \mathcal{X}^K_{L,M}$ is defined by $x_\ell := 1$ for all $\ell \in [1\ldots L]$, while $x_k := -1$ for all $k \in [L+1\ldots K]$. If $\eta_L\,\widetilde{\mu}_L = \eta_{L+1}\,\widetilde{\mu}_{L+1}$, then the sums (A5) and (A6) have more than one maximizer, but they have the same maximizers, so that $F^{\alpha}_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M},\mu) = F^{\beta}_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M},\mu)$. ◇ Claim 1

Now, fix $\boldsymbol{\lambda} \in \mathbb{R}^{\mathcal{K}}_+$, and consider the judgement context $\boldsymbol{\mathcal{C}} := (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}^K_{L,M}, \mathcal{X}^K_{L,M})$. Let $\boldsymbol{\kappa} := \boldsymbol{\lambda}^{1/\alpha}$. Then

$$F^{\alpha}(\boldsymbol{\mathcal{C}},\mu) \underset{(*)}{=} F^{\alpha}_{\boldsymbol{\lambda}}(\mathcal{X}^K_{L,M},\mu) \underset{(\dagger)}{=} F^{1}_{\boldsymbol{\kappa}}(\mathcal{X}^K_{L,M},\mu), \qquad \text{for all } \mu \in \Delta(\mathcal{X}^K_{L,M}). \tag{A7}$$

Here, $(*)$ is obtained by comparing equations (7) and (A3), while $(\dagger)$ follows from Claim 1. Now, $F^1_{\boldsymbol{\kappa}}$ is just the ($\boldsymbol{\kappa}$-weighted) median rule; thus, $F^1_{\boldsymbol{\kappa}}$ satisfies Reinforcement on $\mathcal{X}^K_{L,M}$. Thus, $F^{\alpha}_{\boldsymbol{\lambda}}$ also satisfies Reinforcement on $\mathcal{X}^K_{L,M}$. This proves the first assertion of Proposition 4.

If $\boldsymbol{\lambda} = (1,1,\ldots,1)$, then $\boldsymbol{\kappa} = \boldsymbol{\lambda}$. Thus $F^1_{\boldsymbol{\kappa}} = F^1_{\boldsymbol{\lambda}}$. Thus, statement (A7) implies that $F^{\alpha}(\boldsymbol{\mathcal{C}},\mu) = F^{1}(\boldsymbol{\mathcal{C}},\mu)$ for all $\mu \in \Delta(\mathcal{X}^K_{L,M})$ —in other words, F^{α} itself is the median rule on $\boldsymbol{\mathcal{C}}$. This proves the second assertion of Proposition 4.

On the other hand, if $\boldsymbol{\lambda} \neq (1,1,\ldots,1)$, then $\boldsymbol{\kappa} \neq \boldsymbol{\lambda}$. Thus, $F^1_{\boldsymbol{\lambda}}$ and $F^1_{\boldsymbol{\kappa}}$ will not always agree on $\mathcal{X}^K_{L,M}$. (Discrepancies between these two rules can be constructed using reasoning similar to the proof of Claim 1.) Thus, statement (A7) implies that F^{α} is *not* the median rule on $\boldsymbol{\mathcal{C}}$. This proves the third assertion of Proposition 4. □

We will frequently use the following result.

Lemma A.3 *Any additive majority rule is SME on any judgement context.*

Proof: See Corollary 5.2 from Nehring and Pivato (2018). □

Appendix B: Proofs of the main results

In the proofs of our main results, we will require the judgement context and the rule to satisfy one of two hypotheses. We will say that a judgement context $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ has ***balanced weights*** if for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ with only two issues $i, j \in \mathcal{K}$ such that $x_i \neq y_j$ and $x_j \neq y_j$ (so that $d(\mathbf{x},\mathbf{y}) = 2$), we have $\lambda_i = \lambda_j$. If F is a judgement aggregation rule, then we will say that F is ***compatible*** with $\boldsymbol{\mathcal{C}}$ if there is some $\mu \in \Delta(\mathcal{Y})$ such that $F(\boldsymbol{\mathcal{C}},\mu) = \{\mathbf{x},\mathbf{y}\}$ for some $\mathbf{x},\mathbf{y} \in \mathcal{X}$ with $d(\mathbf{x},\mathbf{y}) \geq 3$.

Note that any *unweighted* judgement context automatically has balanced weights. Thus, we will later invoke balanced weights to prove Theorem 1. On the other hand, we will invoke compatibility to prove Theorem B.2. Indeed, Theorems 1 and B.2 are both consequences of the following, more general result.

Theorem B.1 *Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a weighted judgement context where $\mathcal{Y}$ is thick, and let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be an additive majority rule. Suppose that either $\boldsymbol{\mathcal{C}}$ has balanced weights or F is compatible with $\boldsymbol{\mathcal{C}}$. Then F satisfies* Continuity *and* Reinforcement *if and only if it is the median rule* (8).

Before proving Theorem B.1, we introduce some notation. Let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be any additive majority rule, and let $\mathcal{C} := \text{conv}(\mathcal{Y})$. For any $\mathbf{c} \in \mathcal{C}$, there exists some $\mu \in \Delta(\mathcal{Y})$ such that $\widetilde{\boldsymbol{\mu}} = \mathbf{c}$. We then define $F(\mathbf{c}) := F(\boldsymbol{\mathcal{C}}, \mu)$. By inspection of defining formulae (4) and (7), it is clear that this definition is independent of the choice of μ. Thus, we can define a correspondence $F : \mathcal{C} \rightrightarrows \mathcal{X}$. We will make use of this convention frequently in what follows. For any $\mathbf{x} \in \mathcal{X}$, define

$$\mathcal{C}^F_{\mathbf{x}} \quad := \quad \{\mathbf{c} \in \mathcal{C}\ ;\ \mathbf{x} \in F(\mathbf{c})\} \quad \text{and} \quad {}^o\mathcal{C}^F_{\mathbf{x}} \quad := \quad \{\mathbf{c} \in \mathcal{C}\ ;\ F(\mathbf{c}) = \{\mathbf{x}\}\}. \tag{B1}$$

Let ${}^o\mathcal{C}$ be the topological interior of $\mathcal{C}$ as a subset of $\mathbb{R}^{\mathcal{K}}$. (Note that ${}^o\mathcal{C} \neq \emptyset$ because $\mathcal{Y}$ is thick.) For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, we define

$$\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \quad := \quad \mathcal{C}^F_{\mathbf{x}} \cap \mathcal{C}^F_{\mathbf{y}} \quad = \quad \{\mathbf{c} \in \mathcal{C}\ ;\ \mathbf{x}, \mathbf{y} \in F(\mathbf{c})\}, \tag{B2}$$

$$\text{and} \quad {}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \quad := \quad \{\mathbf{c} \in {}^o\mathcal{C}\ ;\ F(\mathbf{c}) = \{\mathbf{x}, \mathbf{y}\}\}. \tag{B3}$$

Let $\mathcal{K}(\mathbf{x}, \mathbf{y}) := \{k \in \mathcal{K};\ x_k \neq y_k\}$. Thus, for any $\mu \in \Delta(\mathcal{Y})$, and any gain function $\phi : [-1, 1] \longrightarrow {}^*\mathbb{R}$, we have

$$(\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) \quad = \quad \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, (x_k - y_k)\, \phi(\widetilde{\mu}_k), \tag{B4}$$

because $(x_k - y_k) = 0$ for all $k \in \mathcal{K} \setminus \mathcal{K}(\mathbf{x}, \mathbf{y})$. In particular, if $\widetilde{\boldsymbol{\mu}} \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ —that is, if $\{\mathbf{x}, \mathbf{y}\} \subseteq F(\mathcal{X}, \mu)$ —then the sum (B4) must be zero.

Proof of Theorem B.1. It is easy to verify that the median rule satisfies the Reinforcement and Continuity. It remains to verify the converse. So, let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be an additive majority rule with gain function $\phi : [-1, 1] \longrightarrow {}^*\mathbb{R}$.

Claim 1: *Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, and let $\mathbf{z} := \frac{1}{2}(\mathbf{x} + \mathbf{y})$. Then $\mathbf{z} \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$.*

Proof: Observe that $z_k = x_k = y_k$ for all $k \in \mathcal{K} \setminus \mathcal{K}(\mathbf{x}, \mathbf{y})$, while $z_k = 0$ for all $k \in \mathcal{K}(\mathbf{x}, \mathbf{y})$. Let $\epsilon > 0$ and let $\mathbf{z}^\epsilon := \epsilon\, \mathbf{x} + (1 - \epsilon)\, \mathbf{z}$. Then $\text{sign}(z^\epsilon_k) = x_k$ for all $k \in \mathcal{K}$; thus, $F(\mathbf{z}^\epsilon) = \{\mathbf{x}\}$ by supermajority efficiency (because of Lemma A.3). Likewise, if $\mathbf{z}^{-\epsilon} := \epsilon\, \mathbf{y} + (1 - \epsilon)\, \mathbf{z}$, then $\text{sign}(z^{-\epsilon}_k) = y_k$ for all $k \in \mathcal{K}$; thus, $F(\mathbf{z}^{-\epsilon}) = \{\mathbf{y}\}$ by supermajority efficiency. However, clearly $\lim_{\epsilon \to 0} \mathbf{z}^\epsilon = \mathbf{z} = \lim_{\epsilon \to 0} \mathbf{z}^{-\epsilon}$. Thus, $\mathbf{z}$ is a cluster point of both $\mathcal{C}^F_{\mathbf{x}}$ and $\mathcal{C}^F_{\mathbf{y}}$, so $\mathbf{z} \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ because F is satisfies Continuity. ◇ Claim 1

Claim 2: *Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, and let $\mathbf{b} \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. For all $s \in [0, 1)$, we have $s\,\mathbf{b} + (1-s)\mathbf{x} \in {}^o\mathcal{C}^F_{\mathbf{x}}$ and $s\,\mathbf{b} + (1 - s)\mathbf{y} \in {}^o\mathcal{C}^F_{\mathbf{y}}$.*

Proof: Lemma A.3 says that F is supermajority efficient (SME); thus, $F(\mathbf{x}) = \{\mathbf{x}\}$. By hypothesis, $\{\mathbf{x}, \mathbf{y}\} \subseteq F(\mathbf{b})$. Thus, for all $s \in [0, 1)$ we have $F(s\,\mathbf{b} + (1-s)\mathbf{x}) = F(\mathbf{x}) \cap F(\mathbf{b}) = \{\mathbf{x}\}$, by Reinforcement. Thus, $s\,\mathbf{b} + (1-s)\mathbf{x} \in {}^o\mathcal{C}^F_{\mathbf{x}}$. By the same argument, $s\,\mathbf{b} + (1-s)\mathbf{y} \in {}^o\mathcal{C}^F_{\mathbf{y}}$. ◇ `Claim 2`

Claim 3: *For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if there exists $\mu \in \Delta(\mathcal{Y})$ with $F(\boldsymbol{\mathcal{C}}, \mu) = \{\mathbf{x}, \mathbf{y}\}$, then ${}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ is nonempty, and it is a relatively open subset of some affine hyperplane in $\mathbb{R}^{\mathcal{K}}$.*

Proof: Let $K := |\mathcal{K}|$. For all $\mathbf{x} \in \mathcal{X}$, Reinforcement implies that $\mathcal{C}^F_{\mathbf{x}}$ is a convex subset of $\mathcal{C}$. Continuity implies that $\mathcal{C}^F_{\mathbf{x}}$ is closed. Thus, $\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ is closed and convex, because $\mathcal{B}^F_{\mathbf{x},\mathbf{y}} = \mathcal{C}^F_{\mathbf{x}} \cap \mathcal{C}^F_{\mathbf{y}}$. But Claim 2 implies that $\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ has empty interior; thus, it is a convex subset of $\mathbb{R}^{\mathcal{K}}$ with dimension at most $K - 1$.

SME implies that $F(\mathbf{x}) = \{\mathbf{x}\}$. Thus, Continuity implies that $F(\mathbf{c}) = \{\mathbf{x}\}$ for all $\mathbf{c} \in \mathcal{C}$ in some open ball around $\mathbf{x}$. Thus, $\mathcal{C}^F_{\mathbf{x}}$ contains an open ball around $\mathbf{x}$. Thus, $\mathcal{C}^F_{\mathbf{x}}$ itself has nonempty interior, so it is a closed, convex set of dimension K.

Now, let $\mathbf{b} := \widetilde{\mu}$. Then $F(\mathbf{b}) = \{\mathbf{x}, \mathbf{y}\}$. Thus, Continuity yields some $\epsilon > 0$ such that, if $\mathcal{D}$ is the ball of radius ϵ around $\mathbf{b}$ in $\mathcal{C}$, then $F(\mathbf{d}) \subseteq \{\mathbf{x}, \mathbf{y}\}$ for all $\mathbf{d} \in \mathcal{D}$. Define the convex sets:

$$\mathcal{D}_{\mathbf{x}} \ := \ \mathcal{D} \setminus \mathcal{C}^F_{\mathbf{y}}, \quad \mathcal{D}_{\mathbf{y}} \ := \ \mathcal{D} \setminus \mathcal{C}^F_{\mathbf{x}}, \quad \text{and} \quad \mathcal{D}_0 \ := \ \mathcal{D} \cap \mathcal{B}^F_{\mathbf{x},\mathbf{y}}.$$

Thus, $\mathcal{D}_{\mathbf{x}} \subset {}^o\mathcal{C}^F_{\mathbf{x}}$, and $\mathcal{D}_{\mathbf{y}} \subset {}^o\mathcal{C}^F_{\mathbf{y}}$, and $\mathcal{D} = \mathcal{D}_{\mathbf{x}} \sqcup \mathcal{D}_0 \sqcup \mathcal{D}_{\mathbf{y}}$. Also, $\mathcal{D}_{\mathbf{x}}$ and $\mathcal{D}_{\mathbf{y}}$ are nonempty, because Claim 2 implies that the line segment from $\mathbf{b}$ to $\mathbf{x}$ lies in $\mathcal{D}_{\mathbf{x}}$, while the line segment from $\mathbf{b}$ to $\mathbf{y}$ lies in $\mathcal{D}_{\mathbf{y}}$.

Claim 3A: *The set $\mathcal{D} \setminus \mathcal{D}_0$ is path-disconnected.*

Proof: Note that $\mathcal{D} \setminus \mathcal{D}_0 = \mathcal{D}_{\mathbf{x}} \sqcup \mathcal{D}_{\mathbf{y}}$. Let $\mathbf{d}_{\mathbf{x}} \in \mathcal{D}_{\mathbf{x}}$ and $\mathbf{d}_{\mathbf{y}} \in \mathcal{D}_{\mathbf{y}}$. Let $\mathcal{P} \subset \mathcal{D}$ be any path from $\mathbf{d}_{\mathbf{x}}$ to $\mathbf{d}_{\mathbf{y}}$. Then $F(\mathbf{p}) \subseteq \{\mathbf{x}, \mathbf{y}\}$ for all $\mathbf{p} \in \mathcal{P}$. Continuity implies that $F(\mathbf{p}) = \{\mathbf{x}\}$ for all points $\mathbf{p} \in \mathcal{P}$ close to $\mathbf{d}_{\mathbf{x}}$ and $F(\mathbf{p}) = \{\mathbf{y}\}$ for all points $\mathbf{p} \in \mathcal{P}$ close to $\mathbf{d}_{\mathbf{y}}$. Thus, Continuity yields some $\mathbf{p}_0 \in \mathcal{P}$ such that $F(\mathbf{p}_0) = \{\mathbf{x}, \mathbf{y}\}$; thus $\mathbf{p}_0 \in \mathcal{D}_0$. Thus, $\mathbf{d}_{\mathbf{x}}$ and $\mathbf{d}_{\mathbf{y}}$ lie in different path components of $\mathcal{D} \setminus \mathcal{D}_0$. ▽ `Claim 3A`

Claim 3A implies that $\mathcal{D}_0$ meets the interior of $\mathcal{D}$, which means it meets ${}^o\mathcal{C}$ (because $\text{int}(\mathcal{D}) \subseteq {}^o\mathcal{C}$). But clearly $\mathcal{D}_0 \cap {}^o\mathcal{C} \subseteq {}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$; thus, we deduce that ${}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \neq \emptyset$, as claimed.

Furthermore, $\mathcal{D}_0$ is a convex subset of $\mathcal{D}$ of dimension at most $K - 1$, which cuts the ball $\mathcal{D}$ into at least two disconnected pieces (by Claim 3A). The only way this can happen is if $\mathcal{D}_0 = \mathcal{D} \cap \mathbb{H}$ for some affine hyperplane $\mathbb{H} \subset \mathbb{R}^{\mathcal{K}}$. Now, $\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ is a convex subset of $\mathbb{R}^{\mathcal{K}}$, and we have just found an open ball $\mathcal{D} \subset \mathbb{R}^{\mathcal{K}}$ such that $\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \cap \mathcal{D} \subset \mathbb{H}$; thus, $\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subset \mathbb{H}$. Thus, ${}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subset \mathbb{H}$.

Finally, for any point $\mathbf{b} \in {}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$, we can repeat the above construction to obtain an $\mathbb{H}$-relatively open neighbourhood $\mathcal{D}_0$ around $\mathbf{b}$ in ${}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$; thus, ${}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ is a relatively open subset of $\mathbb{H}$. ◇ `Claim 3`

Claim 4: (a) *For any $\mathbf{x} \in \mathcal{X}$, $\mathcal{C}^F_{\mathbf{x}}$ is a closed, convex polyhedron in $\mathbb{R}^{\mathcal{K}}$, and $(\partial \mathcal{C}^F_{\mathbf{x}}) \cap \mathcal{C} = \bigcup_{\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{x}\}} \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$.*

(b) *For all* $\mathbf{y} \in \mathcal{X}$, *if* ${}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \neq \emptyset$, *then it is contained a codimension-1 face of* $(\partial\mathcal{C}^F_{\mathbf{x}}) \cap \mathcal{C}$.

(c) *Conversely, for each each codimension-1 face* $\mathcal{F}$ *of* $(\partial\mathcal{C}^F_{\mathbf{x}}) \cap \mathcal{C}$, *there is some* $\mathbf{y} \in \mathcal{X}$ *such that* $\emptyset \neq {}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subseteq \mathcal{F}$.

Proof: (a) $\mathcal{C}^F_{\mathbf{x}}$ is closed by Continuity, and convex by Reinforcement. For any $\mathbf{y} \in \mathcal{X}$, $\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subset \partial\mathcal{C}^F_{\mathbf{x}}$ by Claim 2. This proves that $(\partial\mathcal{C}^F_{\mathbf{x}}) \cap \mathcal{C} \supseteq \bigcup_{\mathbf{y}\in\mathcal{X}\setminus\{\mathbf{x}\}} \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. To see the opposite inclusion, let $\mathbf{b} \in (\partial\mathcal{C}^F_{\mathbf{x}}) \cap \mathcal{C}$. Then $\mathbf{b}$ is a cluster point of $\mathcal{C}^F_{\mathbf{x}}$, but $\mathbf{b}$ is also a cluster point of $\mathcal{C}^F_{\mathbf{y}}$ for some $\mathbf{y} \neq \mathbf{x}$. Thus, $\mathbf{b} \in \mathcal{C}^F_{\mathbf{x}}$ and $\mathbf{b} \in \mathcal{C}^F_{\mathbf{y}}$ by Continuity, so $\mathbf{b} \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. Thus, $(\partial\mathcal{C}^F_{\mathbf{x}}) \cap \mathcal{C} \subseteq \displaystyle\bigcup_{\mathbf{y}\in\mathcal{X}\setminus\{\mathbf{x}\}} \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$.

(b) If ${}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \neq \emptyset$, then Claim 3 says it is a nonempty open subset of some hyperplane, hence of codimension 1, hence contained in some codimension-1 face of $\partial\mathcal{C}^F_{\mathbf{x}}$.

(c) Let $\mathcal{F}$ be a codimension-1 face of $(\partial\mathcal{C}^F_{\mathbf{x}}) \cap \mathcal{C}$; we claim that ${}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subseteq \mathcal{F}$ for some $\mathbf{y} \in \mathcal{X}$. To see this, first note that $\mathcal{F} \subseteq \bigcup_{\mathbf{y}\in\mathcal{X}\setminus\{\mathbf{x}\}} \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$, by part (a). Since this is a finite collection of sets, there must be some $\mathbf{y} \in \mathcal{X}$ such that $\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \cap \mathcal{F}$ is a subset of $\mathcal{F}$ with nonempty relative interior. Call this relative interior set $\mathcal{B}_0$.

We claim that $\mathcal{B}_0 \subset {}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. To see this, consider the set $\mathcal{D}_{\mathbf{y}} := \{s\mathbf{b} + (1-s)\mathbf{y};\ \mathbf{b} \in \mathcal{B}_0$ and $s \in (0,1)\}$. This is an open cone in $\mathbb{R}^{\mathcal{K}}$ with base $\mathcal{B}_0$, and Reinforcement says that $\mathcal{D}_{\mathbf{y}} \subset {}^{\circ}\mathcal{C}^F_{\mathbf{y}}$. Likewise, if we define $\mathcal{D}_{\mathbf{x}} := \{s\mathbf{b} + (1-s)\mathbf{x};\ \ \mathbf{b} \in \mathcal{B}_0$ and $s \in (0,1)\}$, then $\mathcal{D}_{\mathbf{x}}$ is an open cone with base $\mathcal{B}_0$, and Reinforcement says that $\mathcal{D}_{\mathbf{x}} \subset {}^{\circ}\mathcal{C}^F_{\mathbf{x}}$. Note that $\mathcal{D}_{\mathbf{y}}$ and $\mathcal{D}_{\mathbf{x}}$ are disjoint. Since both have $\mathcal{B}_0$ as their base, it follows that the set $\mathcal{O} := \mathcal{D}_{\mathbf{y}} \sqcup \mathcal{B}_0 \sqcup \mathcal{D}_{\mathbf{x}}$ is an open subset of $\mathbb{R}^{\mathcal{K}}$, containing $\mathcal{B}_0$.

Now let $\mathbf{b} \in \mathcal{B}_0$, and suppose by contradiction that $\mathbf{b} \notin {}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. Then $F(\mathbf{b}) \supseteq \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ for some distinct $\mathbf{z} \in \mathcal{X} \setminus \{\mathbf{x}, \mathbf{y}\}$. Thus, if we define $\mathcal{L} := \{s\mathbf{b} + (1-s)\mathbf{z};\ \ s \in (0,1]\}$, then $\mathcal{L}$ is a line-segment with one end at $\mathbf{b}$, and Reinforcement says that $\mathcal{L} \subset {}^{\circ}\mathcal{C}^F_{\mathbf{z}}$. But $\mathcal{O}$ is an open neighbourhood of $\mathbf{b}$, so $\mathcal{L}$ must pass through $\mathcal{O}$ to reach $\mathbf{b}$. Thus, ${}^{\circ}\mathcal{C}^F_{\mathbf{z}} \cap \mathcal{O}$ is nonempty. But since it is an intersection of two open sets, ${}^{\circ}\mathcal{C}^F_{\mathbf{z}} \cap \mathcal{O}$ itself is open; thus, it must contain points in either $\mathcal{D}_{\mathbf{x}}$ or $\mathcal{D}_{\mathbf{y}}$. But this is impossible, because these are subsets of ${}^{\circ}\mathcal{C}^F_{\mathbf{x}}$ and ${}^{\circ}\mathcal{C}^F_{\mathbf{y}}$, which are disjoint from ${}^{\circ}\mathcal{C}^F_{\mathbf{z}}$ by definition. To avoid contradiction, we must have $\mathbf{b} \in {}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$.

From this, it follows that $\mathcal{F} \cap {}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \neq \emptyset$ (because it contains $\mathcal{B}_0$). However, part (b) tells us that ${}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ is entirely contained in some codimension-1 face of $\mathcal{C}$; thus, we must have ${}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subseteq \mathcal{F}$. ◇ `Claim 4`

The strategy of the proof is now as follows. Claim 4(a) tells us that the sets $\{\mathcal{C}^F_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{X}}$ partition $\mathcal{C}$ into closed, convex polyhedra, which overlap only on their boundaries. By a similar argument, the median rule also partitions $\mathcal{C}$ into closed, convex polyhedra $\{\mathcal{C}^{\text{med}}_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{X}}$, which overlap only on their boundaries. We will show that these two partitions are identical. To do this, it suffices to show that the codimension-1 faces of $\mathcal{C}^F_{\mathbf{x}}$ and $\mathcal{C}^{\text{med}}_{\mathbf{x}}$ are the same, for each $\mathbf{x} \in \mathcal{X}$. Claim 4(b,c) tells us that the codimension-1 faces of $\mathcal{C}^F_{\mathbf{x}}$ can be identified with the "boundary" sets ${}^{\circ}\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ for $\mathbf{y} \in \mathcal{X}$; thus, it suffices to show that these boundary sets coincide with those of the median rule.

The boundary face $\partial\mathcal{B}^{\mathrm{med}}_{\mathbf{x},\mathbf{y}}$ lies in the hyperplane $\mathbb{H}_{\mathbf{x},\mathbf{y}} = \{\mathbf{r} \in \mathbb{R}^{\mathcal{K}};\ \mathbf{r} \underset{\boldsymbol{\lambda}}{\bullet} (\mathbf{x}-\mathbf{y}) = 0\}$. We will show that $\partial\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$ also lies in this hyperplane (see Claim 10 below). To do this, we will show that, if we start with a point in $\partial\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$ and perturb it slightly by a vector parallel to $\mathbb{H}_{\mathbf{x},\mathbf{y}}$, then it remains in $\partial\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$. But to perform such a perturbation analysis on points in $\partial\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$, we must perform a corresponding perturbation analysis on the gain function ϕ. To do this, we need some control over "ϕ-increments" of the form $\phi(r+\delta) - \phi(r)$, where $r \in [-1,1]$ and δ is a "small" perturbation. To acheive this, Claims 5 to 9 establish more and more precise control over these ϕ-increments. For all $i \in \mathcal{K}$, let $\mathbf{e}_i := (0,0,\ldots,0,1,0,\ldots,0)$, where the 1 appears in the ith coordinate.

Claim 5: *Let $\mathbf{x},\mathbf{y} \in \mathcal{X}$, with $d(\mathbf{x},\mathbf{y}) \geq 2$ and let $\mathbf{b} \in \partial\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$. Let $i,j \in \mathcal{K}(\mathbf{x},\mathbf{y})$ with $i \neq j$.*

(a) *For any $\delta_i, \delta_j \in \mathbb{R}$,*

$$(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\mathbf{b}+\delta_i\,\mathbf{e}_i+\delta_j\,\mathbf{e}_j) \;=\; \lambda_i\,(x_i-y_i)\,\Big(\phi(b_i+\delta_i)-\phi(b_i)\Big)+\lambda_j\,(x_j-y_j)\,\Big(\phi(b_j+\delta_j)-\phi(b_j)\Big).$$

(b) *Let $r := b_i$ and $s = b_j$. There exists $\epsilon_{rs} > 0$ and a unique constant $c_{rs} > 0$ (which is determined by $\mathbf{x}$, $\mathbf{y}$, i, and j) such that, for any $\epsilon \in (-\epsilon_{rs}, \epsilon_{rs})$, we have $\phi(r+\epsilon) - \phi(r) = \frac{\lambda_j}{\lambda_i}\Big(\phi(s + c_{rs}\,\epsilon) - \phi(s)\Big)$.*

Proof: (a) $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\mathbf{b}+\delta_i\,\mathbf{e}_i+\delta_j\,\mathbf{e}_j)$

$$\begin{aligned}
&\underset{(\diamond)}{=}\; \sum_{k\in\mathcal{K}(\mathbf{x},\mathbf{y})\setminus\{i,j\}} \lambda_k\,(x_k-y_k)\phi(b_k) + \lambda_i\,(x_i-y_i)\phi(b_i+\delta_i) + \lambda_j\,(x_j-y_j)\phi(b_j+\delta_j)\\
&=\; \sum_{k\in\mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\,(x_k-y_k)\phi(b_k) + \lambda_i(x_i-y_i)\Big(\phi(b_i+\delta_i)-\phi(b_i)\Big)\\
&\qquad + \lambda_j\,(x_j-y_j)\Big(\phi(b_j+\delta_j)-\phi(b_j)\Big)\\
&\underset{(\diamond)}{=}\; (\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\mathbf{b}) + \lambda_i\,(x_i-y_i)\,\Big(\phi(b_i+\delta_i)-\phi(b_i)\Big) + \lambda_j\,(x_j-y_j)\,\Big(\phi(b_j+\delta_j)-\phi(b_j)\Big)\\
&\underset{(\dagger)}{=}\; \lambda_i\,(x_i-y_i)\,\Big(\phi(b_i+\delta_i)-\phi(b_i)\Big) + \lambda_j\,(x_j-y_j)\,\Big(\phi(b_j+\delta_j)-\phi(b_j)\Big).
\end{aligned}$$

Here, both $(\diamond)$ are because $(\mathbf{x}-\mathbf{y})_k = (x_k-y_k) = 0$ for all $k \in \mathcal{K}\setminus\mathcal{K}(\mathbf{x},\mathbf{y})$, and $(\dagger)$ is because $\mathbf{b} \in \mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$, so that $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\mathbf{b}) = 0$.

(b) By negating the i coordinate and/or j coordinate of $\mathcal{X}$ and $\mathcal{Y}$ if necessary, we can assume without loss of generality that $x_i = y_j = 1$ and $x_j = y_i = -1$. Claim 3 yields some vector $\mathbf{v} \in \mathbb{R}^{\mathcal{K}}$ and some constant $a \in \mathbb{R}$ such that $\partial\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$ is a relatively open subset of the affine hyperplane $\mathbb{H} := \{\mathbf{r} \in \mathbb{R}^{\mathcal{K}};\ \mathbf{v}\bullet\mathbf{r} = a\}$. Let $c_{rs} := -v_i/v_j$. For any $\epsilon \in \mathbb{R}$, let $\mathbf{b}_\epsilon := \mathbf{b} + \epsilon\,\mathbf{e}_i + c_{rs}\,\epsilon\,\mathbf{e}_j$. Then

$$\mathbf{b}_\epsilon \bullet \mathbf{v} \;=\; \mathbf{b}\bullet\mathbf{v} + \epsilon\,\mathbf{e}_i\bullet\mathbf{v} - \frac{v_i}{v_j}\,\epsilon\,\mathbf{e}_j\bullet\mathbf{v} \;=\; a + \epsilon\,v_i - \frac{v_i}{v_j}\,\epsilon\,v_j \;=\; a. \qquad \text{(B5)}$$

Thus, $\mathbf{b}_\epsilon \in \mathbb{H}$. Thus, if $|\epsilon|$ is small enough, then $\mathbf{b}_\epsilon \in {}^o\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. Thus,

$$0 \quad = \quad (\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\mathbf{b}_\epsilon) \underset{(*)}{=} 2\left[\lambda_i \left(\phi(r+\epsilon)-\phi(r)\right) - \lambda_j \left(\phi(s+c_{rs}\,\epsilon)-\phi(s)\right)\right],$$

where $(*)$ is by setting $\delta_i := \epsilon$ and $\delta_j := c_{rs}\,\epsilon$ in part (a), and noting that $(x_i - y_i) = 2$ while $(x_j - y_j) = -2$. Thus, we conclude that $\phi(r+\epsilon)-\phi(r) = \frac{\lambda_j}{\lambda_i}\left(\phi(s+c_{rs}\,\epsilon)-\phi(s)\right)$, as desired. Finally, observe that $c_{rs} > 0$ and is unique, because ϕ is strictly increasing. $\diamondsuit$ Claim 5

Claim 6: *Let $\mathbf{x},\mathbf{y} \in \mathcal{X}$ be such that $d(\mathbf{x},\mathbf{y}) \geq 3$ and ${}^o\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \neq \emptyset$. Let $\mathbf{b} \in {}^o\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. Let $i,j \in \mathcal{K}(\mathbf{x},\mathbf{y})$ with $i \neq j$. Let $r := b_i$ and $s = b_j$.*

(a) *There is an open interval $\mathcal{T}_r \subseteq [-1,1]$ containing r such that, for all $t \in \mathcal{T}_r$, we have $c_{ts} = c_{rs}$, where c_{rs} and c_{ts} are as in Claim 5(b).*

(b) *For all $t \in \mathcal{T}_r$, there is an open interval $\mathcal{E}_{tr}$ containing 0 such that, for all $\epsilon \in \mathcal{E}_{tr}$, we have $\phi(r+\epsilon)-\phi(r) = \phi(t+\epsilon)-\phi(t)$.*

Proof: (a) Since $d(\mathbf{x},\mathbf{y}) \geq 3$, there is a third coordinate $k \in \mathcal{K}(\mathbf{x},\mathbf{y}) \setminus \{i,j\}$.

Claim 3 yields some vector $\mathbf{v} \in \mathbb{R}^{\mathcal{K}}$ and some constant $a \in \mathbb{R}$ such that ${}^o\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ is a relatively open subset of the affine hyperplane $\mathbb{H} := \{\mathbf{r} \in \mathbb{R}^{\mathcal{K}};\ \mathbf{v} \bullet \mathbf{r} = a\}$. For any $\eta \in \mathbb{R}$, let $\mathbf{b}^\eta := \mathbf{b} + \eta\,\mathbf{e}_i - (v_i/v_k)\,\eta\,\mathbf{e}_k$. Then $\mathbf{b}^\eta \in \mathbb{H}$, by an argument identical to equation (B5). Thus, there is some $\overline{\eta} > 0$ such that $\mathbf{b}^\eta \in {}^o\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ for all $\eta \in (-\overline{\eta},\overline{\eta})$. Let $\mathcal{T}_r := (r-\overline{\eta}, r+\overline{\eta})$. Let $t \in \mathcal{T}_r$, and let $\eta := t - r$. Then $\mathbf{b}^\eta \in {}^o\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$, and by construction we have $b^\eta_i = b_i + \eta = r + \eta = t$, while $b^\eta_j = b_j = s$. Then, repeating the construction in Claim 5(b) using $\mathbf{b}^\eta$ in place of $\mathbf{b}$, we get $\phi(t+\epsilon)-\phi(t) = \frac{\lambda_j}{\lambda_i}\left(\phi(s+c_{ts}\epsilon)-\phi(s)\right)$, for all sufficiently small ϵ, where $c_{ts} = -v_i/v_j$. This works for all $t \in \mathcal{T}_r$. But $r \in \mathcal{T}_r$; thus, in particular $c_{rs} = -v_i/v_j$. Thus, $c_{ts} = c_{rs}$ for all $t \in \mathcal{T}_r$, as claimed.

(b) Let $t \in \mathcal{T}_r$, and let $\epsilon_t := \min\{\epsilon_{rs}, \epsilon_{ts}\}$, where these are defined as in Claim 5(b). Then $\epsilon_t > 0$, and for all $\epsilon \in (-\epsilon_t, \epsilon_t)$, we have

$$\phi(r+\epsilon)-\phi(r) \underset{(*)}{=} \frac{\lambda_j}{\lambda_i}\left(\phi(s+c_{r,s}\,\epsilon)-\phi(s)\right) \underset{(\dagger)}{=} \phi(t+\epsilon)-\phi(t),$$

where $(*)$ is by Claim 5(b), and $(\dagger)$ is by Claim 5(b) and part (a). $\diamondsuit$ Claim 6

Let $\overline{R} := \sup\{b_i;\ \mathbf{b} \in {}^o\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ and $i \in \mathcal{K}(\mathbf{x},\mathbf{y})$ for some $\mathbf{x},\mathbf{y} \in \mathcal{X}$ with $d(\mathbf{x},\mathbf{y}) \geq 3\}$. Note that $\overline{R} > 0$ if and only if F and $\boldsymbol{\mathcal{C}}$ are compatible. Thus, Claims 7, 8 and 9 (below) are non-vacuous if $\boldsymbol{\mathcal{C}}$ is compatible with F.

Claim 7: (Assuming compatibility) *For all $r \in \left(0,\overline{R}\right)$, there is an open interval $\mathcal{T}_r$ containing r, and for all $t \in \mathcal{T}_r$, there is an open interval $\mathcal{E}_{tr}$ containing 0, such that, for any $\epsilon \in \mathcal{E}_{tr}$, we have $\phi(t+\epsilon)-\phi(t) = \phi(r+\epsilon)-\phi(r)$.*

Proof: Since $r \leq \overline{R}$, there exist $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ with $d(\mathbf{x}, \mathbf{y}) \geq 3$ and some $\mathbf{b} \in {}^{o}\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$ and $i \in \mathcal{K}(\mathbf{x}, \mathbf{y})$ such that $r \leq b_i \leq \overline{R}$. Let $s = r/b_i$; thus, $s \in [0, 1]$ and $r = s\, b_i$. Let $\mathbf{z} := \frac{1}{2}(\mathbf{x} + \mathbf{y})$, and then let $\mathbf{b}^s := s\,\mathbf{b} + (1-s)\mathbf{z}$. Then $F(\mathbf{b}^s) = F(\mathbf{b}) \cap F(\mathbf{z}) = \{\mathbf{x}, \mathbf{y}\}$, because F satisfies Reinforcement and $F(\mathbf{z}) \supseteq \{\mathbf{x}, \mathbf{y}\}$ by Claim 1. Thus, $\mathbf{b}^s \in {}^{o}\mathcal{B}^{F}_{\mathbf{x},\mathbf{y}}$. Note that $b^s_i = s\, b_i = r$. Thus, Claim 6 yields some neighbourhood $\mathcal{T}_r$ around r, and for all $t \in \mathcal{T}_r$, an open interval $\mathcal{E}_{tr}$ containing 0, such that, for all $t \in \mathcal{T}_r$ and $\epsilon \in \mathcal{E}_{tr}$ we have $\phi(t + \epsilon) - \phi(t) = \phi(r + \epsilon) - \phi(r)$. ◇ Claim 7

Claim 8: (Assuming compatibility) *For all $r, s \in \left(0, \overline{R}\right)$, there exists $\overline{\epsilon} = \overline{\epsilon}(r, s) > 0$ containing 0, such that, for any $\epsilon \in (-\overline{\epsilon}, \overline{\epsilon})$, we have $\phi(s + \epsilon) - \phi(s) = \phi(r + \epsilon) - \phi(r)$.*

Proof: Without loss of generality, suppose $r < s$. For all $q \in [r, s]$, let $\mathcal{T}_q$ be as in Claim 7. The family $\{\mathcal{T}_q\}_{q \in [r,s]}$ is an open cover of the compact set $[r, s]$, so it has a finite subcover, say $\{\mathcal{T}_{q_0}, \ldots, \mathcal{T}_{q_N}\}$, where $r = q_0 \leq q_1 < q_2 < \cdots < q_N \leq s = q_N$. By dropping to a subsequence of $\{q_0, q_1, \ldots, q_N\}$ if necessary, we can ensure that $q_n \in \mathcal{T}_{q_{n-1}}$ for all $n \in [1 \ldots N]$ (because these intervals cover $[r, s]$). For all $n \in [1 \ldots N]$, let $\mathcal{E}_{q_n, q_{n-1}}$ be the open interval around 0 defined in Claim 7. Let $\mathcal{E} := \mathcal{E}_{q_1, q_0} \cap \mathcal{E}_{q_2, q_1} \cap \cdots \cap \mathcal{E}_{q_N, q_{N-1}}$; then $\mathcal{E}$ is an open interval containing 0, so there is some $\overline{\epsilon} > 0$ such that $(-\overline{\epsilon}, \overline{\epsilon}) \subseteq \mathcal{E}$. For all $\epsilon \in (-\overline{\epsilon}, \overline{\epsilon})$, we have

$$\phi(q_N + \epsilon) - \phi(q_N) \;=\; \phi(q_{N-1} + \epsilon) - \phi(q_{N-1}) \;=\; \cdots \;=\; \phi(q_1 + \epsilon) - \phi(q_1) \;=\; \phi(q_0 + \epsilon) - \phi(q_0),$$

where each equality is an invocation of Claim 7. In other words, $\phi(s + \epsilon) - \phi(s) = \phi(r + \epsilon) - \phi(r)$, as claimed. ◇ Claim 8

Claim 9: (Assuming compatibility) *For all $r \in \left(0, \overline{R}\right)$, there is some $\delta = \delta(r) > 0$ such that for any $q \in \mathbb{Q} \cap [-1, 1]$, we have $\phi(r + q\,\delta) - \phi(r) = q \cdot [\phi(r + \delta) - \phi(s)]$.*

Proof: Find $\delta > 0$ such that $0 < r - \delta < r + \delta < \overline{R}$. Thus, if we define $\mathcal{S} := [r - \delta, r + \delta]$, then $\mathcal{S} \subset (0, \overline{R})$. Thus, for all $s \in \mathcal{S}$, Claim 8 yields some $\epsilon_s := \overline{\epsilon}(r, s) > 0$ such that $\phi(s + \epsilon) - \phi(s) = \phi(r + \epsilon) - \phi(r)$ for all $\epsilon \in (-\epsilon_s, \epsilon_s)$. It is easily verified that the function $\mathcal{S} \ni s \mapsto \epsilon_s \in \mathbb{R}_+$ is continuous. The interval $\mathcal{S}$ is compact. Thus, there exists some $\epsilon' > 0$ such that $\epsilon_s \geq \epsilon'$ for all $s \in \mathcal{S}$.

Now, let $M_0 \in \mathbb{N}$ be large enough that $\delta/M_0 < \epsilon'$. For any rational number $q \in [-1, 1]$, we can write $q = N/M$ for some $N \in [-M \ldots M]$ and some $M \in \mathbb{N}$ with $M \geq M_0$. Thus, if we define $\epsilon := \delta/M$, then $\epsilon < \epsilon'$, and $q\,\delta = N\,\epsilon$, and we have

$$\begin{aligned} \phi(r + q\,\delta) - \phi(r) \;&=\; \phi(r + N\,\epsilon) - \phi(r) \;=\; \sum_{n=1}^{N} \Big(\phi(r + n\,\epsilon) - \phi(r + (n-1)\,\epsilon)\Big) \\ &\underset{(*)}{=} \sum_{n=1}^{N} \Big(\phi(r + \epsilon) - \phi(r)\Big) \;=\; N \cdot \Big(\phi(r + \epsilon) - \phi(r)\Big) \\ &=\; N \cdot \left(\phi\left(r + \frac{1}{M}\delta\right) - \phi(r)\right), \end{aligned} \tag{B6}$$

where (∗) is by N applications of Claim 8 (which applies because for all $n \in [1 \ldots N]$, we have $s := r + (n-1)\,\epsilon \in \mathcal{S}$ and $\epsilon < \epsilon' \leq \epsilon_s$). In particular, if $q = 1$ (so that $N = M$), then (B6) yields

$$\phi(r+\delta) - \phi(r) \quad = \quad M \cdot \left(\phi\left(r + \frac{\delta}{M}\right) - \phi(r)\right),$$

which means that

$$\phi\left(r + \frac{1}{M}\delta\right) - \phi(r) \quad = \quad \frac{1}{M}\left(\phi(r+\delta) - \phi(r)\right). \qquad \text{(B7)}$$

For any $q = N/M \in [-1, 1]$, if we substitute (B7) into (B6), we get

$$\phi(r + q\,\delta) - \phi(r) \quad = \quad \frac{N}{M} \cdot \Big(\phi\,(r+\delta) - \phi(r)\Big) \quad = \quad q \cdot \Big(\phi\,(r+\delta) - \phi(r)\Big),$$

as desired. ◇ Claim 9

Claim 10: *Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ be distinct, and let $\mathbb{H}_{\mathbf{x},\mathbf{y}} := \{\mathbf{r} \in \mathbb{R}^{\mathcal{K}};\ (\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \mathbf{r} = 0\}$. If ${}^{o}\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \neq \emptyset$, then ${}^{o}\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subset \mathbb{H}_{\mathbf{x},\mathbf{y}}$.*

Proof: First suppose $d(\mathbf{x}, \mathbf{y}) = 1$, and let $\mathcal{K}(\mathbf{x}, \mathbf{y}) = \{i\}$. Then $\mathbb{H}_{\mathbf{x},\mathbf{y}} = \{\mathbf{r} \in \mathbb{R}^{\mathcal{K}};\ r_i = 0\}$. By negating the ith coordinate of $\mathcal{X}$ and $\mathcal{Y}$ if necessary, we can assume without loss of generality that $x_i = 1$ and $y_i = -1$. Then for any $\mu \in \Delta(\mathcal{Y})$, if $\widetilde{\mu}_i > 0$ then $\mathbf{y} \notin \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$ (because $\gamma_{\mathbf{x},\mu}(q) \geq \gamma_{\mathbf{y},\mu}(q)$ for all $q \in [0, 1]$, and $\gamma_{\mathbf{x},\mu}(q) \geq \gamma_{\mathbf{y},\mu}(q) + \lambda_i$ for all $q \in [0, \widetilde{\mu}_i]$.) Likewise, if $\widetilde{\mu}_i < 0$ then $\mathbf{x} \notin \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$. Thus, if $\{\mathbf{x}, \mathbf{y}\} \subseteq \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$, then we must have $\widetilde{\mu}_i = 0$. Since F is SME, it follows that $\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subset \mathbb{H}_{\mathbf{x},\mathbf{y}}$.

Now suppose $d(\mathbf{x}, \mathbf{y}) \geq 2$. There are now two cases: Either $\boldsymbol{\mathcal{C}}$ is compatible with F, or it is *not* compatible with F, but has balanced weights.

Case A. Suppose $\boldsymbol{\mathcal{C}}$ is not compatible with F, but has balanced weights. If $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, and ${}^{o}\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \neq \emptyset$, then we must have $d(\mathbf{x}, \mathbf{y}) \leq 2$ (because $\boldsymbol{\mathcal{C}}$ is not compatible with F). We have already dealt with the case $d(\mathbf{x}, \mathbf{y}) = 1$, so suppose $d(\mathbf{x}, \mathbf{y}) = 2$. Let $\mathcal{K}(\mathbf{x}, \mathbf{y}) = \{i, j\}$. By balanced weights, we must have $\lambda_i = \lambda_j$. For simplicity, suppose $\lambda_i = \lambda_j = 1$. By negating the ith and jth coordinate of $\mathcal{X}$ and $\mathcal{Y}$ if necessary, we can assume without loss of generality that $x_i = y_j = 1$ and $y_i = x_j = -1$. Thus, $\mathbb{H}_{\mathbf{x},\mathbf{y}} = \{\mathbf{r} \in \mathbb{R}^{\mathcal{K}};\ r_i = r_j\}$. Now, for any $\mu \in \Delta(\mathcal{Y})$, if $\widetilde{\mu}_i > \widetilde{\mu}_j$ then $\mathbf{y} \notin \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$ (because $\gamma_{\mathbf{x},\mu}(q) \geq \gamma_{\mathbf{y},\mu}(q)$ for all $q \in [0, 1]$, and $\gamma_{\mathbf{x},\mu}(q) \geq \gamma_{\mathbf{y},\mu}(q) + 1$ for all $q \in (\widetilde{\mu}_j, \widetilde{\mu}_i]$.) Likewise, if $\widetilde{\mu}_i < \widetilde{\mu}_j$ then $\mathbf{x} \notin \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$. Thus, if $\{\mathbf{x}, \mathbf{y}\} \subseteq \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$, then we must have $\widetilde{\mu}_i = \widetilde{\mu}_j$. Since F is SME, it follows that $\mathcal{B}^F_{\mathbf{x},\mathbf{y}} \subset \mathbb{H}_{\mathbf{x},\mathbf{y}}$.

Case B. Suppose F and $\boldsymbol{\mathcal{C}}$ are compatible. Then $\overline{R} > 0$. Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, and suppose $d(\mathbf{x}, \mathbf{y}) \geq 2$. Let $\mathcal{J} := \mathcal{K}(\mathbf{x}, \mathbf{y})$ and let $\mathcal{L} := \mathcal{K} \setminus \mathcal{J}$. By negating the i and/or j coordinates of $\mathcal{X}$ and $\mathcal{Y}$ if necessary, we can assume without loss of generality that $x_j = 1$ and $y_j = -1$ for all $j \in \mathcal{J}$. Let $\mathbf{z} := \frac{1}{2}(\mathbf{x} + \mathbf{y})$. Let $\mathbf{b} \in {}^{o}\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. For all $u \in [0, 1]$, let $\mathbf{b}^u := u\,\mathbf{b} + (1-u)\mathbf{z}$. Then $\mathbf{b}^u \in {}^{o}\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ by Claim 1 and Reinforcement, as explained above. If u is small enough, then we have $b^u_i \in (0, \overline{R})$ for all $i \in \mathcal{K}(\mathbf{x}, \mathbf{y})$.

Claim 10A: *For any distinct $i, j \in \mathcal{K}(\mathbf{x}, \mathbf{y})$, there exists $\epsilon > 0$ such that, $\mathbf{b}^u + \epsilon\,(\lambda_j\,\mathbf{e}_i - \lambda_i\,\mathbf{e}_j) \in {}^{o}\!\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$.*

Proof: Without loss of generality, suppose $\lambda_i \geq \lambda_j$. For any $\delta > 0$ and $\theta \in [0, 1]$, let $\mathbf{b}^\theta_\delta := \mathbf{b}^u + \theta\delta\,\mathbf{e}_i - \delta\,\mathbf{e}_j$. Since $F(\mathbf{b}^u) = \{\mathbf{x}, \mathbf{y}\}$ and F satisfies Continuity, there exists some $\overline{\epsilon} > 0$ such that, for all $\epsilon_1, \epsilon_2 \in (-\overline{\epsilon}, \overline{\epsilon})$, we have $F(\mathbf{b}^u + \epsilon_1\,\mathbf{e}_i + \epsilon_2\,\mathbf{e}_j) \subseteq \{\mathbf{x}, \mathbf{y}\}$. In particular, for any $\delta \in (0, \overline{\epsilon})$, and $\theta \in [0, 1]$, we have $F(\mathbf{b}^\theta_\delta) \subseteq \{\mathbf{x}, \mathbf{y}\}$. Without loss of generality, suppose $\overline{\epsilon} \leq \overline{\epsilon}(b^u_i, b^u_j)$, where $\overline{\epsilon}(b^u_i, b^u_j)$ is as in Claim 8. Let $\delta(b^u_i)$ be defined as in Claim 9, and find some small enough $q \in \mathbb{Q} \cap [0, 1]$ such that, if $\delta := q\,\delta(b^u_i)$, then $\delta \in (0, \overline{\epsilon})$.

Let $\theta := \frac{\lambda_j}{\lambda_i}$. Then $\theta \in [0, 1]$. We will show that $\mathbf{b}^\theta_\delta \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. If we knew that ϕ was linear, then we could deduce that that $\phi(b^u_i + \theta\delta) - \phi(b^u_i) = \theta \cdot [\phi(b^u_i + \delta) - \phi(b^u_i)]$, and from here, use Claims 8 and 5(a) to obtain $(\mathbf{x} - \mathbf{y}) \underset{\lambda}{\bullet} \phi(\mathbf{b}^\theta_\delta) = 0$ and hence $F(\mathbf{b}^\theta_\delta) = \{\mathbf{x}, \mathbf{y}\}$. . But we *don't* know that ϕ is linear; instead, Claim 9 only tells us that ϕ is "locally $\mathbb{Q}$-linear". Thus, we must approximate θ with rational numbers.

Let $\{q_n\}_{n=1}^\infty$ be a decreasing sequence in $\mathbb{Q} \cap [0, 1]$ with $\lim_{n\to\infty} q_n = \theta$. For all $n \in \mathbb{N}$, we have

$$\begin{aligned} \phi(b^u_i + q_n\delta) - \phi(b^u_i) \;&\underset{(*)}{=}\; q_n\,[\phi(b^u_i + \delta) - \phi(b^u_i)] \;\underset{(\dagger)}{=}\; q_n\left[\phi(b^u_j) - \phi(b^u_j - \delta)\right] \\ &\underset{(\diamond)}{>}\; \frac{\lambda_j}{\lambda_i}\left[\phi(b^u_j) - \phi(b^u_j - \delta)\right], \qquad \text{and thus,} \\ \lambda_i\,[\phi(b^u_i + q_n\delta) - \phi(b^u_i)] \;&>\; \lambda_j\left[\phi(b^u_j) - \phi(b^u_j - \delta)\right], \end{aligned} \tag{B8}$$

where $(*)$ is by setting $r := b^u_i$ in Claim 9, $(\dagger)$ is by setting $r := b^u_i$, $s := b^u_j - \delta$, and $\epsilon := \delta$ in Claim 8, and $(\diamond)$ is because $q_n > \theta = \frac{\lambda_j}{\lambda_i}$.

Since $(x_i - y_i) = (x_j - y_j) = 2$, the inequality (B8) yields

$$\lambda_i\,(x_i - y_i)\,\left[\phi(b^u_i + q_n\delta) - \phi(b^u_i)\right] + \lambda_j\,(x_j - y_j)\,\left[\phi(b^u_j - \delta) - \phi(b^u_j)\right] \quad > \quad 0. \tag{B9}$$

Thus, Claim 5(a) yields $(\mathbf{x} - \mathbf{y}) \underset{\lambda}{\bullet} \phi(\mathbf{b}^{q_n}_\delta) > 0$. Thus, $\mathbf{y} \notin F(\mathbf{b}^{q_n}_\delta)$. But we have already noted that $\emptyset \neq F(\mathbf{b}^{q_n}_\delta) \subseteq \{\mathbf{x}, \mathbf{y}\}$. Thus, we must have $F(\mathbf{b}^{q_n}_\delta) = \{\mathbf{x}\}$. But $\lim_{n\to\infty} q_n = \theta$, so $\lim_{n\to\infty} \mathbf{b}^{q_n}_\delta = \mathbf{b}^\theta_\delta$. Thus, Continuity implies that $\mathbf{x} \in F(\mathbf{b}^\theta_\delta)$ also.

Now let $\{q_n\}_{n=1}^\infty$ be an increasing sequence in $\mathbb{Q}\cap[0, 1]$ with $\lim_{n\to\infty} q_n = \theta$. For all $n \in \mathbb{N}$, we obtain $\lambda_i\,(x_i - y_i)\,[\phi(b^u_i + q_n\delta) - \phi(b^u_i)] + \lambda_j\,(x_j - y_j)\,\left[\phi(b^u_j - \delta) - \phi(b^u_j)\right] < 0$, by an argument similar to inequality (B9). Thus, Claim 5(a) yields $(\mathbf{x} - \mathbf{y}) \underset{\lambda}{\bullet} \phi(\mathbf{b}^{q_n}_\delta) < 0$. Thus, by an argument similar to the previous paragraph, we get $F(\mathbf{b}^{q_n}_\delta) = \{\mathbf{y}\}$, for all $n \in \mathbb{N}$. But $\lim_{n\to\infty} q_n = \theta$, so $\lim_{n\to\infty} \mathbf{b}^{q_n}_\delta = \mathbf{b}^\theta_\delta$. Thus, Continuity implies that $\mathbf{y} \in F(\mathbf{b}^\theta_\delta)$ also.

Combining these observations, we deduce that $\{\mathbf{x}, \mathbf{y}\} \subseteq F(\mathbf{b}^\theta_\delta)$. But we have already noted that $F(\mathbf{b}^\theta_\delta) \subseteq \{\mathbf{x}, \mathbf{y}\}$. Thus, $F(\mathbf{b}^\theta_\delta) \subseteq \{\mathbf{x}, \mathbf{y}\}$ —that is, $\mathbf{b}^\theta_\delta \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. Now define $\epsilon := \delta/\lambda_i$; then $\mathbf{b}^\theta_\delta = \mathbf{b}^u + \epsilon\,(\lambda_j\,\mathbf{e}_i - \lambda_i\,\mathbf{e}_j)$, which proves the claim.
$\bigtriangledown$ `Claim 10A`

Claim 10B: *Let $\ell \in \mathcal{L}$. If $\epsilon > 0$ is small enough, then $\mathbf{b}^u + \epsilon\,\mathbf{e}_\ell \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$.*

Proof: Let $\epsilon > 0$. If ϵ is small enough, then Continuity implies that $\emptyset \neq F(\mathbf{b}^u + \epsilon\,\mathbf{e}_\ell) \subseteq \{\mathbf{x}, \mathbf{y}\}$ (because $F(\mathbf{b}^u) = \{\mathbf{x}, \mathbf{y}\}$). But $(\mathbf{x} - \mathbf{y}) \underset{\lambda}{\bullet} \phi(\mathbf{b}^u + \epsilon\,\mathbf{e}_\ell) = (\mathbf{x} - \mathbf{y}) \underset{\lambda}{\bullet} \phi(\mathbf{b}^u) = 0$ (because $\phi(\mathbf{b}^u + \epsilon\,\mathbf{e}_\ell)_j = \phi(b_j)$ for all $j \in \mathcal{J}$, while $(\mathbf{x} - \mathbf{y})_k = 0$ for all $k \in \mathcal{L}$). Thus, we must have $F(\mathbf{b}^u + \epsilon\,\mathbf{e}_\ell) = \{\mathbf{x}, \mathbf{y}\}$; hence $\mathbf{b}^u + \epsilon\,\mathbf{e}_\ell \in \mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. $\bigtriangledown$ `Claim 10B`

Claim 3 says that ${}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$ is a relatively open subset of some affine hyperplane $\mathbb{H}$. Claims 10A and 10B imply that $\mathbb{H}$ is parallel to all vectors in the set $\{(\lambda_j\,\mathbf{e}_i - \lambda_i\,\mathbf{e}_j);\ i,j \in \mathcal{J}\} \cup \{\mathbf{e}_\ell;\ \ell \in \mathcal{L}\}$. But the hyperplane $\mathbb{H}_{\mathbf{x},\mathbf{y}}$ is spanned by this set. Thus, $\mathbb{H}$ is parallel to $\mathbb{H}_{\mathbf{x},\mathbf{y}}$. Let $\mathbf{z} = (\mathbf{x}+\mathbf{y})/2$. Then $\mathbf{z} \in \mathbb{H}_{\mathbf{x},\mathbf{y}}$. But Claim 1 and Reinforcement imply that $\mathbf{z}$ is a cluster point of ${}^o\mathcal{B}^F_{\mathbf{x},\mathbf{y}}$. Thus, $\mathbb{H} = \mathbb{H}_{\mathbf{x},\mathbf{y}}$. ◇ Claim 10

For any $\mathbf{x} \in \mathcal{X}$, let $\mathcal{C}^{\mathrm{med}}_{\mathbf{x}} := \{\mathbf{c} \in \mathcal{C};\ \ \mathbf{x} \in \mathrm{Median}\,(\mathcal{X}, \mathbf{c})\}$. Then $\mathcal{C}^{\mathrm{med}}_{\mathbf{x}}$ is a convex polyhedron whose supporting hyperplanes are the sets $\mathbb{H}_{\mathbf{x},\mathbf{y}}$ (from Claim 10) for all $\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{y}\}$, along with the supporting hyperplanes of $\mathcal{C}$ itself. Claims 4 and 10 show that every one of these supporting hyperplanes is also a supporting hyperplane of the convex polyhedron $\mathcal{C}^F_{\mathbf{x}}$. Thus, $\mathcal{C}^F_{\mathbf{x}} \subseteq \mathcal{C}^{\mathrm{med}}_{\mathbf{x}}$ for all $\mathbf{x} \in \mathcal{X}$. However, the systems $\{\mathcal{C}^F_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{X}}$ and $\{\mathcal{C}^{\mathrm{med}}_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{X}}$ are each partitions of $\mathcal{C}$ into closed convex polyhedra which meet only along their boundaries. Thus, they must be identical. Thus, F is the median rule. □

Proof of Theorem 1. It is easy to verify that the median rule (3) satisfies ESME, Continuity, and Reinforcement; we must verify the converse. So, suppose F is a rule satisfying these axioms. Proposition 1 says that F is an additive majority rule, because it satisfies ESME and Continuity. Any unweighted judgement context obviously has balanced weights. Thus, if F also satisfies Reinforcement, then Theorem B.1 says it is the median rule. □

Theorem 2 is a consequence of a more general result, involving a more complicated structural condition. Let F be a judgement aggregation rule on the context $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$, and let $\Theta \subset \Delta(\mathcal{Y})$ be a collection of profiles. Let $\mathcal{V}^F_\theta := \{\mathbf{x} \in \mathcal{X};\ \ \exists\ \mu \in \Theta$ such that $F(\mu) = \{\mathbf{x}\}\}$. For any $\mathbf{v}, \mathbf{w} \in \mathcal{V}^F_\theta$, we write $\mathbf{v} \overset{F}{\sim} \mathbf{w}$ if $d(\mathbf{v},\mathbf{w}) \leq 2$ and there is some $\mu \in \Theta$ with $F(\mu) = \{\mathbf{v}, \mathbf{w}\}$. Thus, $(\mathcal{V}^F_\Theta, \overset{F}{\sim})$ is a graph. This graph is ***path-connected*** if, for any $\mathbf{u}, \mathbf{w} \in \mathcal{V}^F_\theta$, there is some path $\mathbf{u} \overset{F}{\sim} \mathbf{v}_1 \overset{F}{\sim} \mathbf{v}_2 \overset{F}{\sim} \cdots \overset{F}{\sim} \mathbf{v}_N \overset{F}{\sim} \mathbf{w}$ connecting them in $\mathcal{V}^F_\theta$. We will say that the judgement context $\boldsymbol{\mathcal{C}}$ is ***frangible*** if for *any* additive majority rule F satisfying Continuity, there exists an open, connected subset $\Theta \subset \Delta(\mathcal{Y})$ such that the graph $(\mathcal{V}^F_\Theta, \overset{F}{\sim})$ is *not* path-connected. Theorem 2 is a corollary of the next result.

Theorem B.2 *Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a frangible weighted judgement context such that $\mathcal{Y}$ is thick, and let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be a judgement aggregation rule. Then F satisfies* WESME, Continuity, *and* Reinforcement *if and only if F is the median rule* (8).

Proof: Suppose $\boldsymbol{\mathcal{C}}$ is frangible and $\mathcal{Y}$ is thick, and F is an additive majority rule satisfying Continuity and Reinforcement. It suffices to show that $\boldsymbol{\mathcal{C}}$ and F are compatible.

Since $\boldsymbol{\mathcal{C}}$ is frangible, there is some open, connected $\Theta \subseteq \Delta(\mathcal{Y})$ and some $\mathbf{v}, \mathbf{w} \in \mathcal{V}^F_\Theta$ in different $\overset{F}{\sim}$-connected components. Let $\widetilde{\Theta} := \{\widetilde{\boldsymbol{\mu}};\ \ \mu \in \Theta\}$; this is an open, connected subset of $\mathbb{R}^{\mathcal{K}}$ (because the function $\Delta(\mathcal{Y}) \ni \mu \mapsto \widetilde{\boldsymbol{\mu}} \in \mathbb{R}^{\mathcal{K}}$ is open and continuous). Let $\mathcal{C} = \mathrm{conv}(\mathcal{Y})$; then $\widetilde{\Theta} \subseteq \mathcal{C}$. As explained after Claim 4 in the proof of Theorem B.1, the sets $\{\mathcal{C}^F_{\mathbf{x}}\}_{\mathbf{x}\in\mathcal{X}}$ partition $\mathcal{C}$ into closed, convex polyhedra, which overlap only on their boundaries. Now, $\widetilde{\Theta}$ intersects $\mathcal{C}^F_{\mathbf{v}}$ and $\mathcal{C}^F_{\mathbf{w}}$ (because $\mathbf{v}, \mathbf{w} \in \mathcal{V}^F_\Theta$). Thus, it is possible to construct a continuous path $\alpha : [0,1] \longrightarrow \widetilde{\Theta}$ with $\alpha(0) \in \mathcal{C}_{\mathbf{v}}^F$ and $\alpha(1) \in \mathcal{C}_{\mathbf{w}}^F$, such that for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if α crosses from $\mathcal{C}_{\mathbf{x}}^F$ to $\mathcal{C}_{\mathbf{y}}^F$, then it does so by passing through the codimension-1 face between $\mathcal{C}_{\mathbf{x}}^F$ and $\mathcal{C}_{\mathbf{y}}^F$ —call this face $\mathcal{F}_{\mathbf{x},\mathbf{y}}$.

In fact, we will now show that we can assume without loss of generality that for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if α crosses from $\mathcal{C}_{\mathbf{x}}^F$ to $\mathcal{C}_{\mathbf{y}}^F$, then it does so by passing through the set ${}^o\mathcal{B}_{\mathbf{x},\mathbf{y}}^F$. To see this, let $t \in [0,1]$, and suppose $\alpha(t) \in \mathcal{F}_{\mathbf{x},\mathbf{y}}$ for some $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Claim 4(b,c) in the proof of Theorem B.1 says that ${}^o\mathcal{B}_{\mathbf{x},\mathbf{y}}^F$ is nonempty, and is a subset of $\mathcal{F}_{\mathbf{x},\mathbf{y}}$. So, let $\mathbf{b} \in {}^o\mathcal{B}_{\mathbf{x},\mathbf{y}}^F$, let $\epsilon > 0$, and replace α with the path α' defined:

$$\alpha'(s) \quad := \quad \begin{cases} \alpha(s) & \text{if} \quad s \notin (t-\epsilon, t+\epsilon); \\ (\epsilon - |s-t|)\,\mathbf{b} + (1 - \epsilon + |s-t|)\,\alpha(s) & \text{if} \quad s \in (t-\epsilon, t+\epsilon). \end{cases}$$

Recall that $\widetilde{\Theta}$ is open; thus, if ϵ is small enough, then $\alpha'(s) \in \widetilde{\Theta}$ for all $s \in [0,1]$; furthermore, α' passes through exactly the same polyhedral cells as α, and its passage through all other faces is unchanged. However, $\alpha'(t) = \epsilon\mathbf{b} + (1-\epsilon)\alpha(t)$. Thus, Reinforcement yields $F[\alpha'(t)] = F(\mathbf{b}) \cap F[\alpha(t)] = \{\mathbf{x}, \mathbf{y}\}$, so that $\alpha'(t) \in {}^o\mathcal{B}_{\mathbf{x},\mathbf{y}}^F$, as desired.

Thus, we can construct α such that for all $t \in [0,1]$, either $\alpha(t) \in \mathcal{C}_{\mathbf{x}}^F$ for some $\mathbf{x} \in \mathcal{X}$ (so that $F[\alpha(t)] = \{\mathbf{x}\}$) or $\alpha(t) \in {}^o\mathcal{B}_{\mathbf{x},\mathbf{y}}^F$ for some $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ (so that $F[\alpha(t)] = \{\mathbf{x}, \mathbf{y}\}$).

Now, suppose that $\boldsymbol{\mathcal{C}}$ is *not* compatible with F. Then whenever $\alpha(t) \in {}^o\mathcal{B}_{\mathbf{x},\mathbf{y}}^F$ for some $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, we must have $d(\mathbf{x}, \mathbf{y}) = 2$, and hence, $\mathbf{x} \overset{F}{\sim} \mathbf{y}$. Thus, the function $F \circ \alpha$ defines a path from $\mathbf{v}$ to $\mathbf{w}$ in the graph $(\mathcal{V}_\Theta^F, \overset{F}{\sim})$. But $\mathbf{v}$ and $\mathbf{w}$ are in different connected components of $(\mathcal{V}_\Theta^F, \overset{F}{\sim})$. Contradiction.

To avoid this contradiction, $\boldsymbol{\mathcal{C}}$ must be compatible with F. Then Theorem B.1 says that F is the median rule (8). □

Case (a) of Theorem 2 follows from Theorem B.2 and the next result.

Lemma B.1 *Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. If $\mathcal{X}$ is rugged and $\mathcal{Y}$ is McGarvey, then $\boldsymbol{\mathcal{C}}$ is frangible.*

Proof: Let $\mathbf{z} \in \{\pm 1\}^{\mathcal{K}} \setminus \mathcal{X}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{X}_{\mathbf{z}}$ be such that $d(\mathbf{x}, \mathbf{z}) \neq d(\mathbf{y}, \mathbf{z})$. (By ruggedness, such a $\mathbf{z}$ exists). Let $\Theta := \{\mu \in \Delta(\mathcal{Y});\ \ \mathbf{x}^\mu = \mathbf{z}\} = \{\mu \in \Delta(\mathcal{Y});\ \ \mathrm{sign}(\widetilde{\mu}_k) = z_k$, for all $k \in \mathcal{K}\}$.

Claim 1: *Θ is nonempty, open and convex (hence, connected).*

Proof: Θ is nonempty because $\mathcal{Y}$ is McGarvey. It is defined by a finite system of strict linear inequalities, so it is open and convex. ◇ `Claim 1`

Now let F be any additive majority rule satisfying Continuity. We will show that $(\mathcal{V}_\Theta^F, \overset{F}{\sim})$ is disconnected. First, we need some terminology. A view $\mathbf{x} \in \mathcal{X}$ is *Condorcet admissible* for μ if there does not exist any other $\mathbf{y} \in \mathcal{X}$ such that $y_k\,\widetilde{\mu}_k \geq x_k\,\widetilde{\mu}_k$ for all $k \in \mathcal{K}$ —in other words, there is no view $\mathbf{y} \in \mathcal{X}$ which agrees with the majority in a strictly larger set of issues than those where $\mathbf{x}$ agrees with the majority. Let $\mathrm{Cond}(\mathcal{X}, \mu) \subseteq \mathcal{X}$ be the set of all views that are Condorcet admissible for μ.[21] It is easily verified that supermajority efficiency (for any $\boldsymbol{\lambda}$) implies Condorcet admissibility. Thus, $\mathrm{SME}(\mathcal{X}, \boldsymbol{\lambda}, \mu) \subseteq \mathrm{Cond}(\mathcal{X}, \mu)$.

Claim 2: $\mathcal{V}_\Theta^F = \mathcal{X}_\mathbf{z}$.

Proof: "$\subseteq$ For any $\mu \in \Theta$, we must have $F(\mu) \subseteq \mathrm{SME}(\mathcal{X}, \mu) \subseteq \mathrm{Cond}(\mathcal{X}, \mu)$. But $\mathrm{Cond}(\mathcal{X}, \mu) \subseteq \mathcal{X}_\mathbf{z}$ by Lemma 1.5 of Nehring et al. (2014).

"$\supseteq$" Let $\mathbf{x} \in \mathcal{X}_\mathbf{z}$. Let $\mathcal{J} := \{j \in \mathcal{K};\ x_j = z_j\}$. By negating certain coordinates of $\mathcal{X}$ and $\mathcal{Y}$ if necessary, we can assume without loss of generality that $z_j = x_j = 1$ for all $j \in \mathcal{J}$. Since $\mathcal{Y}$ is McGarvey, there exists some $\mu \in \Delta(\mathcal{Y})$ such that $\widetilde{\mu}_j > 0$ for all $j \in \mathcal{J}$ and $\widetilde{\mu}_k = 0$ for all $k \in \mathcal{K} \setminus \mathcal{J}$ (by Footnote 17). Since $\mathbf{z}$ is near to $\mathbf{x}$, there is no other $\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{x}\}$ such that $y_j = 1$ for all $j \in \mathcal{J}$. Thus, $\mathrm{SME}(\mathcal{X}, \mu) = \{\mathbf{x}\}$. Thus, $F(\mu) = \{\mathbf{x}\}$, because F is supermajority efficient by Lemma A.3.

Let $\theta \in \Theta$ be arbitrary. For all $s \in [0, 1]$, define $\mu^s := s\,\theta + (1-s)\,\mu$. Thus, $\lim_{s \to 0} \mu^s = \mu$. Thus, Continuity implies that $F(\mu^s) = \{\mathbf{x}\}$ for all s sufficiently close to zero. But for all $s > 0$, we have $\mu^s \in \Theta$ (because for all $k \in \mathcal{K}$ we have $\mathrm{sign}(\widetilde{\mu}_k^s) = \mathrm{sign}(s\,\widetilde{\theta}_k + (1-s)\,\widetilde{\mu}_k) = \mathrm{sign}(\widetilde{\theta}_k) = z_k$). Thus, $\mathbf{x} \in \mathcal{V}_\Theta^F$. This argument works for all $\mathbf{x} \in \mathcal{X}_\mathbf{z}$, so $\mathcal{V}_\Theta^F \supseteq \mathcal{X}_\mathbf{z}$, as claimed. ◇ Claim 2

Claim 3: *For any* $\mathbf{v}, \mathbf{w} \in \mathcal{V}_\Theta^F$, *if* $\mathbf{v}$ *and* $\mathbf{w}$ *are in the same* $\overset{F}{\sim}$*-connected component, then* $d(\mathbf{v}, \mathbf{z}) = d(\mathbf{w}, \mathbf{z})$.

Proof: It suffices to prove this in the case when $\mathbf{v} \overset{F}{\sim} \mathbf{w}$; the general case follows by induction on path length. Now, if $\mathbf{v} \overset{F}{\sim} \mathbf{w}$, then $d(\mathbf{v}, \mathbf{w}) = 2$. Thus, $\mathcal{K}(\mathbf{v}, \mathbf{w}) = \{i, j\}$ for some $i, j \in \mathcal{K}$.

Claim 3A: *Either* $i \in \mathcal{K}(\mathbf{v}, \mathbf{z})$ *or* $i \in \mathcal{K}(\mathbf{w}, \mathbf{z})$, *but not both. Likewise, either* $j \in \mathcal{K}(\mathbf{v}, \mathbf{z})$ *or* $j \in \mathcal{K}(\mathbf{w}, \mathbf{z})$, *but not both.*

Proof: Since $v_i = -w_i$, we either have $v_i = -z_i$ or $w_i = -z_i$. But if $v_i = -z_i$, then evidently $w_i = z_i$. This proves the first claim. The second is similar. ▽ Claim 3A

Claim 3B: *Exactly one of* i *or* j *is in* $\mathcal{K}(\mathbf{v}, \mathbf{z})$.

Proof: (by contradiction) If $\{i, j\} \subseteq \mathcal{K}(\mathbf{v}, \mathbf{z})$, then Claim 3A implies $\mathcal{K}(\mathbf{w}, \mathbf{z}) = \mathcal{K}(\mathbf{v}, \mathbf{z}) \setminus \{i, j\}$, which means $\mathcal{K}(\mathbf{w}, \mathbf{z}) \subsetneq \mathcal{K}(\mathbf{v}, \mathbf{z})$, which contradicts the fact that $\mathbf{v}$ is near to $\mathbf{z}$. On the other hand, if $\{i, j\}$ is disjoint from $\mathcal{K}(\mathbf{v}, \mathbf{z})$, then Claim 3A implies $\mathcal{K}(\mathbf{w}, \mathbf{z}) = \mathcal{K}(\mathbf{v}, \mathbf{z}) \sqcup \{i, j\}$, which means $\mathcal{K}(\mathbf{v}, \mathbf{z}) \subsetneq \mathcal{K}(\mathbf{w}, \mathbf{z})$, which contradicts the fact that $\mathbf{w}$ is near to $\mathbf{z}$. ▽ Claim 3B

Without loss of generality, suppose $i \in \mathcal{K}(\mathbf{v}, \mathbf{z})$ and $j \notin \mathcal{K}(\mathbf{v}, \mathbf{z})$. Thus, Claim 3A says $i \notin \mathcal{K}(\mathbf{w}, \mathbf{z})$. But by an argument similar to Claim 3B, exactly one of i or j is in $\mathcal{K}(\mathbf{w}, \mathbf{z})$. Thus, we must have $j \in \mathcal{K}(\mathbf{w}, \mathbf{z})$. At this point, we deduce that $\mathcal{K}(\mathbf{w}, \mathbf{z}) = \{j\} \sqcup \mathcal{K}(\mathbf{v}, \mathbf{z}) \setminus \{i\}$. Thus, $|\mathcal{K}(\mathbf{w}, \mathbf{z})| = |\mathcal{K}(\mathbf{v}, \mathbf{z})|$. In other words, $d(\mathbf{v}, \mathbf{z}) = d(\mathbf{w}, \mathbf{z})$, as claimed. ◇ Claim 3

[21] See Nehring et al. (2014, 2016) and Nehring and Pivato (2014) for an analysis of Condorcet admissibility in judgement aggregation.

By the definition of $\mathbf{z}$, there exist $\mathbf{x}, \mathbf{y} \in \mathcal{X}_{\mathbf{z}}$ such that $d(\mathbf{x}, \mathbf{z}) \neq d(\mathbf{y}, \mathbf{z})$. Claim 2 says that $\mathbf{x}, \mathbf{y} \in \mathcal{V}^F_\Theta$, and Claim 3 implies that they must be in different $\overset{F}{\sim}$-connected components. Thus, $(\mathcal{V}^F_\Theta, \overset{F}{\sim})$ is disconnected, as desired.

This argument works for any additive majority rule (actually, any supermajority efficient rule) satisfying Continuity. Thus, $\mathcal{C}$ is frangible. □

Case (b) of Theorem 2 follows from Theorem B.2 and the next result.

Lemma B.2 *Let $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. If $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$ is distal, then $\mathcal{C}$ is frangible.*

Proof: By hypothesis, there exist some $\mathbf{x}, \mathbf{z} \in \mathcal{X}$ with $\mathbf{x}$ near to $\mathbf{z}$ and $d(\mathbf{x}, \mathbf{z}) \geq 3$. Let $\delta_{\mathbf{x}}$ and $\delta_{\mathbf{z}}$ be the point masses at $\mathbf{x}$ and $\mathbf{z}$, respectively, and let $\mu := \frac{1}{2}(\delta_{\mathbf{x}} + \delta_{\mathbf{z}})$. Let Θ be an open ball of small radius around μ; Θ is obviously open and connected.

Now let F be any additive majority rule satisfying Continuity. We claim that $\mathcal{V}^F_\theta \subseteq \{\mathbf{x}, \mathbf{z}\}$. To see this, let $\mathcal{J} := \mathcal{K}(\mathbf{x}, \mathbf{y})$ and let $\mathcal{L} := \mathcal{K} \setminus \mathcal{J}$. Without loss of generality, suppose $x_\ell = z_\ell = 1$ for all $\ell \in \mathcal{L}$. Then $\widetilde{\mu}_\ell = 1$ for all $\ell \in \mathcal{L}$, while $\widetilde{\mu}_j = 0$ for all $j \in \mathcal{J}$. Fix a weight vector $\boldsymbol{\lambda} \in \mathbb{R}^{\mathcal{K}}_+$, and let $\Lambda = \sum_{\ell \in \mathcal{L}} \lambda_\ell$. Then for any $\mathbf{y} \in \mathcal{X}$ and any $q \in [0, 1]$, we have $\gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{x}}(q) = \gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{z}}(q) = \Lambda \geq \gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{y}}(q)$, with equality if and only if $\mathbf{y}$ is between $\mathbf{x}$ and $\mathbf{z}$. But $\mathbf{x}$ is near to $\mathbf{z}$, so there is no $\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{x}, \mathbf{y}\}$ which is between $\mathbf{x}$ and $\mathbf{z}$. Thus, $\mathrm{SME}\,(\mathcal{X}, \mu) \subseteq \{\mathbf{x}, \mathbf{z}\}$. Thus, Lemma A.3 implies that $F(\mu) \subseteq \{\mathbf{x}, \mathbf{z}\}$. If we make the ball Θ small enough, then Continuity implies that $F(\theta) \subseteq \{\mathbf{x}, \mathbf{z}\}$ for all $\theta \in \Theta$; thus, $\mathcal{V}^F_\theta \subseteq \{\mathbf{x}, \mathbf{z}\}$.

To see that $\mathcal{V}^F_\theta = \{\mathbf{x}, \mathbf{z}\}$, define $\mu_n := \frac{n+1}{2n}\delta_{\mathbf{x}} + \frac{n-1}{2n}\delta_{\mathbf{z}}$ for all $n \in \mathbb{N}$. Then by an argument similar to the previous paragraph, supermajority efficiency implies that $F_n(\mu_n) = \{\mathbf{x}\}$ for all $n \in \mathbb{N}$. If n is large enough, then $\mu_n \in \Theta$; thus, $\mathbf{x} \in \mathcal{V}^F_\Theta$. By an identical argument (defining $\mu_n := \frac{n-1}{2n}\delta_{\mathbf{x}} + \frac{n+1}{2n}\delta_{\mathbf{z}}$ for all $n \in \mathbb{N}$), we obtain $\mathbf{z} \in \mathcal{V}^F_\Theta$. Thus, $\mathcal{V}^F_\theta = \{\mathbf{x}, \mathbf{z}\}$, as claimed. But $d(\mathbf{x}, \mathbf{z}) \geq 3$; thus, $(\mathcal{V}^F_\theta, \sim)$ is not path-connected, and thus, $\mathcal{C}$ is frangible. □

Proof of Theorem 2. It is easy to verify that the weighted median rule (8) satisfies WESME, Continuity, and Reinforcement; we must verify the converse. If F is a rule satisfying WESME and Continuity, then Proposition 3 says that F is an additive majority rule. Suppose F also satisfies Reinforcement. If $\mathcal{X}$ is rugged and $\mathcal{Y}$ is McGarvey, then Lemma B.1 says $\mathcal{C}$ is frangible. On the other hand, if $\mathcal{X}$ is distal and $\mathcal{Y}$ is thick, then Lemma B.2 says $\mathcal{C}$ is frangible. Either way, Theorem B.2 implies that F is the median rule. □

References

Balinski, M., Laraki, R., 2010. Majority Judgment. MIT Press.

Brams, S. J., Fishburn, P. C., 1983. Approval voting. Birkhäuser Boston, Mass.

Day, W. H., McMorris, F. R., 2003. Axiomatic Consensus Theory in Group Choice and Biomathematics. SIAM, Philadelphia.

Dhillon, A., Mertens, J.-F., 1999. Relative utilitarianism. Econometrica 67, 471–498.

Dietrich, F., 2014. Scoring rules for judgment aggregation. Soc. Choice Welf. 42 (4), 873–911.

Emerson, P., 2016. From Majority Rule to Inclusive Politics – Electing a Power-Sharing Coalition. Springer.

Guilbaud, G.-T., Octobre-Décembre 1952. Les théories de l'intérêt général et le problème logique de l'aggrégation. Economie Appliquée V (4), 501–551.

Hanson, R., 2013. Shall we vote on values, but bet on beliefs? Journal of Political Philosophy 21, 151–178.

Kemeny, J. G., Fall 1959. Math without numbers. Daedalus 88, 571–591.

Kornhauser, L., Sager, L., 1986. Unpacking the court. Yale Law Journal 96, 82–117.

Lang, S., Pigozzi, G., Slavkovik, M., van der Torre, L., 2011. Judgement aggregation rules based on minimization. In: Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge. ACM, New York, NY, pp. 238–246.

List, C., Pettit, P., 2002. Aggregating sets of judgements: an impossibility result. Economics and Philosophy 18, 89–110.

List, C., Puppe, C., 2009. Judgement aggregation: a survey. In: Oxford handbook of rational and social choice. Oxford University Press, Oxford, UK., pp. 457–482.

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.

Miller, A., 2013. A model of community standards. Journal of Economic Theory 148 (6), 2696–2705.

Miller, M. K., Osherson, D., 2009. Methods for distance-based judgment aggregation. Social Choice and Welfare 32 (4), 575–601.

Mongin, P., 2012. The doctrinal paradox, the discursive dilemma, and logical aggregation theory. Theory and Decision 73 (3), 315–355.

Monjardet, B., 2008. "mathématique sociale" and mathematics. a case study: Condorcet's effect and medians. Electronic Journal for the History of Probability and Statistics 4, 1–26.

Myerson, R. B., 1995. Axiomatic derivation of scoring rules without the ordering assumption. Soc. Choice Welf. 12 (1), 59–74.

Nehring, K., Pivato, M., 2011. Incoherent majorities: the McGarvey problem in judgement aggregation. Discrete Applied Mathematics 159, 1488–1507.

Nehring, K., Pivato, M., 2014. How indeterminate is sequential majority voting? A judgement aggregation perspective. In: The mathematics of decisions, elections, and games. Vol. 624 of Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 55–88.

Nehring, K., Pivato, M., 2018. Majority rule in the absence of a majority. (preprint).

Nehring, K., Pivato, M., Puppe, C., 2014. The Condorcet set: Majority voting over interconnected propositions. J.Econ.Theory 151, 268–303.

Nehring, K., Pivato, M., Puppe, C., 2016. Unanimity overruled: Majority voting and the burden of history. Journal of Theoretical Politics 28 (4), 552–597.

Nehring, K., Puppe, C., 2007. The structure of strategy-proof social choice I: General characterization and possibility results on median spaces. J.Econ.Theory 135, 269–305.

Smith, J. H., 1973. Aggregation of preferences with variable electorate. Econometrica 41, 1027–1041.

Young, H. P., 1974. A note on preference aggregation. Econometrica 42, 1129–1131.

Young, H. P., Levenglick, A., 1978. A consistent extension of Condorcet's election principle. SIAM J. Appl. Math. 35 (2), 285–300.