Majority rule in the absence of a majority

Nehring, Klaus and Pivato, Marcus

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

30 January 2018

Majority rule in the absence of a majority*

Klaus Nehring† and Marcus Pivato‡

January 30, 2018

Abstract

Which is the best, impartially most plausible consensus view to serve as the basis of democratic group decision when voters disagree? Assuming that the judgment aggregation problem can be framed as a matter of judging a set of binary propositions ("issues"), we develop a multi-issue majoritarian approach based on the criterion of *supermajority efficiency* (SME). SME reflects the idea that smaller supermajorities must yield to larger supermajorities so as to obtain better supported, more plausible group judgments. As it is based on a partial ordering, SME delivers unique outcomes only in special cases. In general, one needs to make cardinal, not just ordinal, tradeoffs between different supermajorities. Hence we axiomatically characterize the class of *additive majority rules*, whose (generically unique) outcome can be interpreted as the "on balance most plausible" consensus judgment.

1 Introduction

1.1 Preference aggregation vs. judgement aggregation

The dominant paradigm of decision-making group in groups by voting is preference aggregation. In preference aggregation, inputs —(votes) are understood to express what is best for each voter personally. The crux of preference aggregation is a conflict of interest. The normative question is thus how to best resolve those conflicts of interest in a fair and efficient manner.

An alternative paradigm of decision-making in groups by voting is *judgement aggregation* (JA). In JA, inputs (votes) concern judgements on the best course of action for

*Earlier versions of the paper have been presented at the 2010 Meeting of Society for Social Choice and Welfare (Moscow), the 2011 Workshop on New Developments in Judgement Aggregation and Voting (Freudenstadt), the University of Montreal (2011), the Paris School of Economics (2012), ESSLLI Duesseldorf (2013), DTEA (Paris 2014), and ETH Zurich (2016). We are grateful to the participants at these presentations for their valuable suggestions.

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

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

the group as a collective agent. Inputs are views about what the group should do, about which decision is right. These judgements may be about the best choice directly, or may concern the basis for such choice in views about what is right as a matter of, say justice or law, or in beliefs about matters of uncertain fact. The crux of judgement aggregation is *disagreement*; the output (group choice) aims at a *disagreement resolution* by means of a consensus. A good consensus is one that has most merit in light of the input judgements. The normative standard is thus, broadly speaking, "epistemic" rather than "ethical".

In this paper, we adopt the less common, and, in our view, still underdeveloped judgement aggregation perspective. Our aim is to develop a broadly applicable "*multi-issue majoritarian*" (MIM) approach whose intellectual ancestry can be tracked back to Condorcet (1785) via Guilbaud (1952). While much of the JA literature has focused on impossibility results, we offer a *possibility* result by providing a general axiomatic characterization of a (novel) class of aggregation rules called *additive majority rules* (AMR). This class encompasses the widely-studied class of *median rules*. See the end of the introduction for the connection to the literature, and section 2 for an informal outline of the basic ideas and axioms underlying our results.

While we do not pin down here a uniquely optimal majoritarian aggregation rule, we obtain a parsimoniously characterized class; this compares favorably to the veritable "zoo" of majoritarian rules in the standard voting literature. We will develop the motivation and content of the majoritarian approach further in the outline section of the paper. In the remainder of the introduction, we will flesh out and motivate the broader judgement aggregation perspective a bit more, as it is not very familiar or even well-understood. Indeed, by providing a well-structured possibility result that is applicable across a range of different settings, we hope to contribute to the broader conceptual development of judgement aggregation as well.

The JA perspective often arises naturally from the institutional context of the voting problem. For example, *committees* are often asked to come up with a decision that is effective or binding for a larger polity. Such committees include corporate boards, central banks, juries of various kinds, and multi-member courts, including supreme courts. "Aggregation" here is about making best use of the committee members' judgemental inputs, rather than about figuring out how best to satisfy their personal preferences.[1]

But the JA perspective is pertinent not only to committees with delegated decision power, but also to *polities* with sovereign decision power. Regarding decisions for public office, one can argue that there is an important distinction between voting on personal preference and voting on a more impersonal view of the merit or quality of the candidate. The merit-oriented perspective was indeed the stance of Lull (1299), Cusanus (1434) and Condorcet (1785) in their pioneering contributions to social choice theory; it has been powerfully developed more recently by Balinski and Laraki (2010). In political referenda, the preference aggregation perspective is sometimes the natural one (for instance in decisions on local public goods), but frequently the natural emphasis is on the perceived merit of a

[1] Of course, making judgements does not presuppose disinterestedness. It will often, if not always, be associated with a preference for obtaining a group judgement and choice that is aligned with one's own judgement. "Strategic manipulability" of votes is thus a concern for judgement aggregation as it is for preference aggregation.

decision, as in referenda, say, on same-sex marriage, to give an example of current interest. In many situations, such as decisions on taxation and redistribution, both perspectives will be pertinent and illuminating.

The relative prevalence of the two perspectives is quite different in different disciplines. In economics, the preference aggregation perspective dominates almost entirely. By contrast, a broader view of group decision by voting is frequently taken in political science and political philosophy, and the JA perspective is much more prevalent, as exemplified by the work of authors such as Waldron (1999), Christiano (2008), Estlund (2008), List and Pettit (2011), and Schwartzberg (2015) .

Much of the more formal literature on JA themes is devoted to variants of the Condorcet Jury Theorem (CJT), which asserts the ability of majorities made of minimally competent voters to discover the truth. Its appeal to the notion of truth in the context of voting and its reliance on statistical models (often with strong assumptions) has made the CJT approach somewhat controversial (see Section 2.2.2); by contrast, the approach here, while broadly epistemic, does not rely on statistical arguments.

1.2 The range of judgement aggregation

Standard voting theory from a preference aggregation perspective has a standard format: its input is a set of linear or weak orders, and its output is either a social welfare order, or the set of "best" (i.e. "social welfare maximizing" or "fairest") choices. By contrast, JA comes in a great variety of formats and structures. We shall mention and briefly describe some of them, but there are many others. The MIM approach is applicable to all of them, sometimes with more, sometimes less degrees of freedom, sometimes with few (if any) evident alternatives, sometimes with strong competitors.

- The most direct JA counterpart to the standard preference aggregation model assumes that choice is derived from a *comparative* evaluation of "merit" or "choiceworthiness" of the alternatives. Each voter submits a *ranking* of the alternatives, expressing her view of comparative merit. A JA rule then determines a "most plausible" consensus ranking as the basis for choice of the group. Thus, ranking aggregation is formally identical to *Arrovian preference aggregation*, but it concerns *rankings* rather than preference orders. (See Example 4.1 below for details.)

- Alternatively, the evaluation of merit could be *non-comparative*, as in grading based on a commonly understood scale. Such non-comparative evaluations are common in athletic or artistic competitions. Although the "grading" of political candidates for office is not common, it is conceptually meaningful. Indeed, it has been proposed and thoroughly elaborated, both theoretically and empirically, in the recent magisterial work by Balinski and Laraki (2010) on "Majority Judgement".

- In many instances, the proximate task of judgement aggregation is to aggregate "beliefs" rather than directly evaluations of overall merit. These beliefs could be quantitative; in particular, they may take the form of subjective probabilities. There is a sizeable literature on the aggregation of probability judgements, most of it involving some form of averaging, and hardly any of it majoritarian (Genest and Zidek, 1986).

- In other settings, such as legal ones, a qualitative model of belief as the *acceptance of propositions* is common and natural. These propositions may concern matters of fact or matters of law. The acceptance model has inspired the *Doctrinal Paradox* due to Kornhauser and Sager (1986) and much of the recent JA literature following it; see in particular List and Pettit (2002). Since the qualitative belief model already comes with a binary issue-frame, the MIM approach seems very natural in this setting.

- There are many other examples. We will elaborate one due to Tangian (2014) in Section 3.1. Tangian's model, which we shall call Virtual Referendum Voting (VRV), attempts to inject elements of direct democracy into an overall political framework of representative democracy. We consider the simplest version in which one "representative" (governing person or governing party) is to be elected on the basis of its views on a number of dichotomous issues that are selected so as to represent the "ideological space". VRV represents a complement of sorts to Balinski and Laraki's "candidate quality" model.

In some settings, there exist serious normative competitors to majoritarian aggregation rules. These are often of an averaging type. As mentioned above, in the literature on aggregating subjective probabilities, averaging-type rules predominate. In aggregating *rankings*, the Borda rule (which averages ranks) is a prominent alternative, while in aggregating *grades*, the *average grade* is an obvious alternative to the median grade. Whether or not these other criteria have real normative appeal, there will often still be strong considerations of a "second-best optimality" nature that favor majoritarian aggregation rules, and in particular, additive majority rules, as sounder and more reliable. Broadly speaking, majoritarian aggregation rules are likely to work better than averaging rules with lower-quality judgemental input, and to incentivize higher-quality input at the same time. Firstly, averaging is highly sensitive to individual judgements. As Galton (1907) observed, the average "gives voting power to cranks in proportion to their crankiness"; Galton thus championed the median, as the least exposed to the influence of unreasonable or erratic opinions. (Note that robustness to cranks is a concern distinctive of judgement aggregation; by contrast, from a preference aggregation perspective, there is no reason to discount unusual preferences.)

Secondly, averaging provides strong and clear-cut *incentives* for any voter (judge) to submit a more extreme judgement. Additive majority rules, by contrast, appear to be significantly more robust against strategic manipulation. In particular, on JA spaces where fully strategy-proof anonymous and neutral aggregation rules exist at all, these rule take the form of issue-by-issue majority voting, and are thus AMRs.[2] AMRs enjoy also "partial strategy-proofness" properties.[3]

[2]Cf. Nehring and Puppe (2007, 2010) building upon Barberà et al. (1997).

[3]See, e.g. Bossert and Sprumont (2014) for the Kemeny rule in the ranking problem.

AMRs are new in the literature, but special cases are not. The median rule, in particular, has been studied extensively. In the context of the ranking problem, the median rule is also known as the *Kemeny rule* (1959), and has been axiomatized in the classic contribution by Young and Levenglick (1978). The median rule has been studied quite extensively for other special classes of judgement spaces as well (Barthélémy and Monjardet, 1981, 1988; Barthélémy and Janowitz, 1991; McMorris et al., 2000).

In addition, two "limiting" cases of AMRs have been studied in the special context of the ranking problem. These are the *Slater rule* (1961), and the *Ranked Pairs* rule proposed by Tideman (1987). While these do not satisfy SME strictly, and hence are not proper AMRs, they posses refinements that *are* proper AMRs, as described in Section 4.

A technically interesting feature of our analysis is the need for representations ranging over the *hyperreal* numbers rather than just real numbers.[4] For example, to ensure SME, the refinements of the Ranked Pairs and Slater rule both require hyperreal-valued representations. We hope that our techniques may be useful in other applications to social choice and decision theory.

The remainder of the paper is organized as follows. Section 2 provides an informal outline of the basic argument, emphasizing intuition and concepts. Section 3 introduces and illustrates the general framework, which goes beyond the standard JA model by allowing an asymmetric treatment of issues via asymmetric (non-uniform) issue weights, and by allowing input and output spaces to differ. Section 4 formally defines the axiom of *supermajority efficiency* and the class of additive majority rules. Section 5 contains our main result: an axiomatic characterization of these rules. Appendix A is a brief formal introduction to hyperreal numbers. Appendices B to F contain the proofs of all the results in the text.

2 Outline of the argument

2.1 The plausibility argument for majority vote

Consider a group faced with a single yes-no decision, which hinges on which alternative is more just, or more in accordance with existing law. Out of the nine members, five favor a "Yes", four a "No". In view of the disagreement, the nine members decide to consult an outside advisor whose sole task is to advise, from her detached, "impartial" point of view, which decision has greater merit on the basis of the anonymized profile of votes alone. The members agree to adopt the advisor's counsel as their *consensus decision*.

The advisor's task is easy, in view of the meagerness of information that characterizes her task. She reasons as follows. I have no basis to distinguish among the nine submitted, anonymized votes. Hence all must count equally. If the vote count was four against four with eight members, there would be equally strong reasons to support both judgements, and I would have no basis to advise for or against one of them. Yet, in point of fact, there

[4] Representations using the hyperreal numbers or other linearly ordered algebraic structures have a long history and decision and social choice theory. See e.g. Halpern (2010,§1) and Pivato (2013a,§7.1) for summaries of this literature. A prominent example in social choice theory is Smith (1973).

is an additional fifth vote in favor of "Yes"; so this tilts the "balance of argument" in favor of Yes, however slightly. We shall say that Yes is "more plausible" or "more plausibly correct", and refer to the above argument as the *Plausibility Argument* for majority vote.

In special cases, judgements admit an independent, fact-like standard of truth; one might then refine the criterion of *plausibility* to the more specific one of *probability.* If judgements are "truth-apt" in this way, the question of optimal judgement aggregation might be turned into a question of optimal statistical inference. This is done in the literature departing from the celebrated Condorcet Jury Theorem; see subsection 2.2.2 below for a more detailed discussion.

However, in this paper, we do not rely on the assumption of truth-aptness, and thus avoid the rather heroic assumptions underlying much of the work in the statistical vein. Even without truth-aptness, judgement is not an entirely subjective affair of mere like and dislike. Some judgement are better-grounded than others; let's call such judgements "plausibility apt". Many judgements in matters of law or of justice seem to be of this kind.[5] Indeed, without some notion of epistemic merit such as "plausibility aptitude", how could one otherwise make sense of any notion of *disagreement* about the right course of action, in contrast to simple *difference in preference*?

2.2 Majoritarianism in complex decisions

How can the logic of this "plausibility argument for majority voting" be extended to more complex decisions? Following a line of intellectual ancestry going back to Condorcet, we will develop a "multi-issue majoritarian" (MIM) approach which assumes that a complex decision/judgement aggregation problem can be sensibly broken down into a set of binary yes-one issues. Formally, we will posit a set $\mathcal{K}$, whose elements represent properties or propositions which can affirmed or denied; we will refer to the elements of $\mathcal{K}$ as ***issues***, and refer to $\mathcal{K}$ itself as a ***frame***. We also posit a set $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$, describing the set of "coherent" or "admissible" views —we call this a ***judgement space***, and refer to elements of $\mathcal{X}$ as ***views***. In many (but not all) cases, a particular frame stands out as the "natural" one. As in most of the judgement aggregation literature, we take the frame as given.[6] Multi-issue Majoritarianism evaluates the overall *plausibility* of a view based on the sign and size of the majority margins on each issue in $\mathcal{K}$. If the issue-wise majorities yield a coherent result (in the so-called "Condorcet consistent" case), then this is the most plausible group judgement; otherwise, additional considerations must be invoked to arrive at a suitably discriminative judgement.

Example 2.1. (*Grading*) Let $(\mathcal{G}, >)$ be a finite, linearly ordered set; we will refer to the elements of $\mathcal{G}$ as ***grades***. For each candidate for some position or office, each voter judges

[5] Obviously, this is rich – and slippery – philosophical terrain. We don't aim to make any philosophically sharp distinctions, but appeal to certain intuitive and common understandings. Sharper conceptual work would likely bear valuable fruit, though.

[6] In the literature, the frame is often referred to as the *agenda* or *property space*. The framing decision itself can be viewed as a separate matter of delegated or group decision. Indeed, Condorcet (1788, p. 140-141) himself devoted some thought to the role of such an issue framing committee.

the candidate's quality by means of a grade in $\mathcal{G}$. Which is the best supported or "most plausible" *consensus* grade for each candidate?

A majoritarian answer can be given by representing the assignment of grades with a frame $\mathcal{K} = \{k_g\}_{g \in \mathcal{G}}$, where the issue k_g represents the statement "the candidate merits grade g or above". The assignment of a grade g can then be identified with a $|\mathcal{G}|$-tuple of yes-no (± 1) judgements, i.e. formally with an element $\mathbf{x}_g$ of the Hamming cube $\{\pm 1\}^{\mathcal{K}}$. For concreteness, let $\mathcal{G} = \{1, 2 \ldots, 7\}$, with the usual ordering; the grade "6" is then represented by the 7-tuple $(+1, +1, +1, +1, +1, +1, -1)$. The judgement space $\mathcal{X}$ consists of the seven views $\{\mathbf{x}_g : g \in \mathcal{G}\}$.

Consider a profile of five voters who submit the grades 2, 3, 6, 6 and 7 for candidate C. For all issues $g \leq 6$, a majority of voters affirms the judgement "C merits grade g or above", while for $g = 7$, a majority of four out of five voters deny the judgement. Hence, the issue-wise majority judgement is the 7-tuple $(+1, +1, +1, +1, +1, +1, -1)$. As this is coherent, corresponding to the assignment of grade "6", the grade 6 is the "impartially most plausible", hence the consensus view. Note that, at this and any other profile, the consensus grade agrees with the *median* of individual grades. It follows that with this issue-framing, MIM coincides with a basic version of Balinski and Laraki's "majority grade".[7] $\diamondsuit$

By always yielding a coherent issue-wise majority view, the grading example is rather special. By contrast, in many other judgement aggregation problems, there will be profiles at which the issue-wise majority view will be incoherent. At such profiles, the issue-wise plausibilities point in different directions, and the best balance of plausibilities must be determined by additional considerations.

Think of the "plausibility" of views at a given profile in terms of a partial order "more plausible than". A baseline —to be refined further —is the partial ordering according to *Condorcet dominance.* A view $\mathbf{x}$ *Condorcet dominates* another view $\mathbf{y}$ if $\mathbf{x}$ agrees with the majority view on all issues on which $\mathbf{y}$ agrees with the majority view, and some others. Equivalently, for any issue in which $\mathbf{x}$ and $\mathbf{y}$ differ, $\mathbf{x}$ agrees with the majority and $\mathbf{y}$ does not. So $\mathbf{x}$ is clearly "issue-wise more plausible" than $\mathbf{y}$, hence $\mathbf{y}$ is ruled out as a possible consensus view. Those views *not* Condorcet dominated by some other coherent view are the *Condorcet admissible* views. A profile is *Condorcet consistent* if has a unique Condorcet admissible view, which holds if and only if the issue-wise majority view is itself coherent (Nehring et al., 2014, Lemma 1.5). In some judgement spaces $(\mathcal{K}, \mathcal{X})$ such as the space of grades in Example 2.1, *all* profiles (with an odd number of voters) are Condorcet consistent.[8] But in all other judgement spaces, there are profiles which admit multiple Condorcet admissible views. Indeed, such indeterminacy arises frequently, and can be quite severe (Nehring et al., 2014, 2016; Nehring and Pivato, 2014). These indeterminacy phenomena roughly parallel and generalize the potentially extreme indeterminacy of the top cycle demonstated in a classical contribution by McKelvey (1986).

[7]Balinski and Laraki refine the aggregation in order to resolve ties among multiple candidates with equal median grades, but such ties need not concern us here.

[8]These are exactly the *median spaces* (Nehring and Puppe, 2010; Bandelt and Barthelemy, 1984).

To select among Condorcet admissible views, some balancing of plausibilities is needed. This can be done by refining the baseline partial ordering by taking into account the *size* of the issue-wise majorities —information that was ignored by Condorcet admissibility, which only exploited their *sign*. The idea is simple: larger majorities make a stronger case for adopting the judgement they support. Such ordinal plausibility tradeoffs lead to the refined criterion of *Supermajority Efficiency* (SME).

To motivate this criterion, consider the classical problem of ranking three alternatives. This ranking problem is naturally framed in terms of three issues, corresponding to the three comparisons a vs. b, b vs. c, and a vs. c. Suppose that, among 12 voters, 5 have the ranking (in downward order) abc, 4 have the ranking bca, and 3 have the ranking cab. At this profile, a net majority of $5+4-3=6$ ranks b above c, a net majority of 4 ranks a above b, and a net majority of 2 ranks c above a. Thus, we have a Condorcet cycle. Condorcet Admissibility discards the rankings cba, acb and bac, which conflict with the majority view on two out of three issues, but it does not select among the remaining three rankings. Yet these are not on par, as they are not equally plausible from an impartial point of view. Compare, for example, the Condorcet admissible rankings abc and cab. Both agree on the ranking of a above b, but they disagree on the other two comparisons. But abc departs from the majority view on an issue with a near tie (on a vs. c with a margin of 2), while the ranking cab departs from the majority view on an issue with a larger majority margin of 6. So from an impartial perspective, abc is more plausible than cab. By a similar argument, abc is also more plausible than bca, and is thus the *most plausible* ranking. Thus, it is the best consensus judgement from a majoritarian perspective.

Note that this argument relies only on ordinal comparisons of the majority margins. This observation motivates our general notion of Supermajority Efficiency, which is designed to give maximal leverage to such ordinal comparisons. In general settings, these comparisons may be more involved than the simple pairwise comparisons in the case of ranking three alternatives. In a nutshell, a view $\mathbf{x}$ *supermajority dominates* another $\mathbf{y}$, if for any given majority margin q, the view $\mathbf{x}$ agrees with supermajorities of size at least q on at least as many issues as $\mathbf{y}$ does, and, for some q, on more. (We will provide a precise definition in Section 3.) A view is *supermajority efficient* if it is not supermajority dominated by any other admissible view.

In the ranking problem with three alternatives, the argument for the profile above generalizes, and it is easy to see that there is (generically) a *unique* SME view. In more complex situations, however, the SME criterion may still leave some indeterminacy. This is the case, for example, in the ranking problem with four or more alternatives. Then there exist profiles and views (that is: rankings) $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x}$ departs from the (incoherent) majority view on a *single* issue with a margin of, say $70\%-30\% = 40\%$, while $\mathbf{y}$ departs from the majority view on *two* issues with margins of, say, $61\% - 39\% = 22\%$ each. Which of these two rankings is more plausible —if either —is not self-evident. A straightforward vote counting criterion suggests $\mathbf{x}$ is superior (since $2 \times 22\% > 40\%$). But the opposite trade-off might be justified as well. For example, one could also argue that the two smaller 22% majority margins for $\mathbf{x}$ are closer to "toss ups", and thus contribute less than proportionately to overall plausibility of a view than the single larger 40% margins for $\mathbf{y}$. Put differently, only large majorities contain a meaningful "signal", while small majorities contain mostly "noise".

Either way, settling on a best consensus in situations of this kind turns on making cardinal trade-offs between super-majorities of different size. The main result of this paper, Theorem 1, shows that if a supermajority-efficient judgement aggregation rule makes these tradeoffs in a systematic and consistent manner across profiles by satisfying an additional axiom (Combination), it must be equal to (or select from) an *additive majority rule*. Such a rule is characterized by a "gain function" ϕ, which maps "majority margins" into "plausibilities". For any profile μ (summarizing the views of the voters), and any admissible view $\mathbf{x}$, the overall plausibility $\Phi(\mathbf{x}, \mu)$ of $\mathbf{x}$ is simply the sum of issue-wise plausibilities. That is,

$$\Phi(\mathbf{x}, \mu) = \sum_{k \in \mathcal{K}} \phi(x_k \, \widetilde{\mu}_k), \tag{2.1}$$

where $\widetilde{\mu}_k$ is the majority margin for $+1$ in issue k, and the views(s) with maximal overall plausibility is chosen as consensus view. The gain function ϕ is increasing, and can be assumed without loss of generality to be odd (i.e. $\phi(-r) = -\phi(r)$ for all r).[9]

This completes the outline of the basic argument of the paper. The formal development in Section 3 generalizes the classical set-up along a number of dimensions. In particular, issues may be weighted unequally, and the admissibility constraints for input and output judgements may differ. But before going into the technical details of this framework, we conclude this Section with a few remarks on the scope and orientation of the overall contribution.

2.2.1 Consensus vs. Compromise

Among additive majority rules, one stands out: the rule in which plausibility is simply proportional to net majority —i.e. in which the gain function ϕ is *linear*. It has been extensively studied in the literature under the name of the *median rule* and, in the special case of the ranking problem, it is the Kemeny rule. Median rules are unique among AMRs in admitting an interpretation as minimizing the *average distance* of the group view to the individual views, where distance among views is given by the number of issues on which the views differ.

By consequence, median rules are distinctively attractive from a compromise perspective. From a compromise perspective, the distance of a group view from an individual view can be interpreted as a measure of the "burden of compromise" imposed on the individual voter; the median rule can then be viewed as minimizing the aggregate burden of compromise. In a companion paper (Nehring and Pivato, 2018), we provide an ax-

[9] For most but not all purposes, plausibility can be taken to be real-valued. However, real-valuedness is not sufficiently general for both material and technical reasons. In general, we must allow ϕ to be hyperreal-valued; hyperreal numbers have just the same arithmetic rules as the reals, but include infinite and infinitesimal numbers besides the reals.

Hyperreals will be seen to be much more accessible and user-friendly for our purposes than their reputation may suggest. In particular, no background in non-standard analysis is required to fully understand the paper. In any case, there is little loss to the big picture if a reader neglects the distinction between reals and hyperreals iomatic foundation of the median rule within the general framework of this paper, but under additional assumptions on the structure of the judgement problem.

2.2.2 "Plausible", not "Probable"

While we have settled here for *plausibility*, much of the literature on epistemic group decision making aims at more, namely *probability.* Following Condorcet's (1785) classic Jury Theorem, this literature assumes that there is an independent standard of "truth" by which the accuracy of judgement decisions can be measured. On the basis of statistical modeling, it attempts to determine the extent and conditions under which majority decisions are likely to be correct, even in cases, in which individual competence is low.[10] This line of reasoning can also be used to justify particular JA rules; a leading example is Young's (1988) reconstruction of Condorcet's derivation of the Kemeny rule as a maximum likelihood estimator. But there are good reasons to doubt whether this quasi-empirical line of reasoning can provide a sound basis for a normative account of group decision as disagreement resolution.[11]

The CJT approach runs into two problems in particular. First, not all matters of reasonable disagreement are truth-apt. Arguably, aiming at "the best" decision does not presuppose that there exists an independent external standard ("truth") as to what this best decision is. For example, in matters of justice and law, disagreeing opinions (and supporting discourses) are arguably all we have.

Second, even if the matter is truth-apt, it is sometimes doubtful whether a sound statistical argument is available. Consider, for example, the aforementioned Condorcet-Young maximum likelihood argument for the median rule. That argument rests on two strong modeling assumptions of stochastic independence: independence across voters and independence across issues. Both of these are questionable, and the inferred maximum-likelihood estimator is highly sensitive to both. Thus, a strand of the recent literature has been devoted to relaxing these independence assumptions (Peleg and Zamir, 2012; Dietrich and Spiekermann, 2013a,b; Pivato, 2013b, 2017). Yet, even if these assumptions are dropped, one is often still left with a difficult statistical inference problem with a tenuous data base. In some cases, the group itself might have to address this problem to arrive at a consensus decision. But this itself could be a significant source of disagreement. Hence, even if statistical arguments have undeniable value as thought experiments, their value as a basis for resolving disagreement in a group may be somewhat limited.

3 Formal Framework

Let $\mathcal{K}$ be a finite set of *issues*. A *view* on $\mathcal{K}$ is an element $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$, where $x_k = 1$ if $\mathbf{x}$ "asserts" proposition k, and $x_k = -1$ if $\mathbf{x}$ "denies" proposition k. A *judgement space* is a subset $\mathcal{X} \subseteq \{\pm 1\}^{\mathcal{K}}$; typically $\mathcal{X}$ is the set of views which are "admissible" or "coherent" according to our interpretation of the elements of $\mathcal{K}$.

[10] See Goodin and Spiekermann (2018) for a book length treatment.

[11] According to McLean and Urken (1995, p. 27-38), Condorcet himself vacillated between "probabilistic" and "straightforward" reasoning. Here we have articulated the "straightforward" line of reasoning.

We have already given two specific examples of judgement spaces: linearly ordered grades and the space of linear orders on a finite set of alternatives. Another example is *truth-functional* aggregation, much studied in the early literature on judgement aggregation inspired by Kornhauser and Sager's (1986) *Doctrinal Paradox*. Here the frame $\mathcal{K}$ consists of a set of logically independent premises $\mathcal{P}$ and a conclusion c that is logically determined by the premises; the judgement space $\mathcal{X} \subset \{\pm 1\}^{\mathcal{K}}$ is the set of logically consistent assignments of truth values. For instance, if $\mathcal{K} = \{p, q, c\}$ and c is the conjunction of p and q, then $\mathcal{X} = \{\mathbf{x} \in \{\pm 1\}^{\mathcal{K}};\ x_c = \min(x_p, x_q)\}$.) More generally, with $\mathcal{K} = \mathcal{P} \cup \{c\}$, the truth-functional judgement spaces are characterized by the condition that, for all $\mathbf{x} \in \mathcal{X}$, $(\mathbf{x}_{\mathcal{P}}, x_c) \in \mathcal{X}$ if and only if $(\mathbf{x}_{\mathcal{P}}, -x_c) \notin \mathcal{X}$.

Other judgement spaces encode social decision problems such as resource allocation, committee selection, and object classification; some of these examples appear later in this paper, while others are discussed by Nehring and Puppe (2007), Nehring and Pivato (2011, 2014), and Nehring et al. (2014, 2016).

The standard judgement aggregation framework typically imposes the same admissiblity/coherence conditions on input and output judgements. But we allow those to differ, denoting the space of admissible *output* views by $\mathcal{X}$ and that of admissible *input* views by $\mathcal{Y}$, with $\mathcal{X} \subseteq \mathcal{Y}$. For the purposes of this paper, this added generality comes essentially for free. It can arise in different ways. In the ranking problem, for example, one may not want to insist on the coherence (transitivity) of individual input comparisons, but require a transitive output ranking to produce a well-defined choice correspondence. In a legal setting, the public justification of a court decision may need to conform to an articulated legal doctrine described by a truth-functional implication, but the reasoning of individual court members may be allowed greater latitude.

Multi-issue majoritarianism must deal with tradeoffs in the alignment of the group view with the majorities on different issues – this, again, is the key motivation for the central axiom of supermajority efficiency. In many judgement problems (such as the ranking problem), it is natural to treat all issue *on par*; at a purely formal level, such parity of issues is underpinned by the structural symmetry of the ranking space under permutations of the alternatives. In other settings, such symmetries may be absent, and there may even be an intrinsic asymmetry among issues. In truth-functional aggregation, for example, there is a fundamental conceptual asymmetry between premises and conclusions. Indeed, in the literature on the Doctrinal Paradox, a major theme has been the question whether the aggregation should be driven by aggregation of premises first, with the conclusion derivative, or by aggregation of the conclusion first. These can be viewed as extreme stances along a continuum in which premises and conclusions receive different *weights*.

We are thus lead to generalize the notion of a *judgement space* $(\mathcal{K}, \mathcal{X})$ to that of a *judgement context* as a quadruple $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$, where $\mathcal{K}$ is a finite set of issues, $\boldsymbol{\lambda} = (\lambda_k)_{k \in \mathcal{K}} \in \mathbb{R}_+^{\mathcal{K}}$ is a positive *weight vector*, and $\mathcal{X} \subseteq \mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}}$.

To focus on the case of issue parity, we will frequently drop reference to the weight vector $\boldsymbol{\lambda}$, and refer to an *unweighted judgement context* as a triple $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$. Formally, unweighted judgement contexts can be identified with judgement contexts with unit weights, i.e. for which $\lambda_k = 1$ for all $k \in \mathcal{K}$.

Let $\mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}}$ be a judgement space, and let $\Delta(\mathcal{Y})$ be the set of all functions μ :

$\mathcal{Y}\longrightarrow\mathbb{R}_+$ such that $\sum_{\mathbf{y}\in\mathcal{Y}}\mu(\mathbf{y})=1$. An element $\mu\in\Delta(\mathcal{Y})$ is called a ***profile***, and describes a population of (weighted) voters; for each $\mathbf{y}\in\mathcal{Y}$, the value of $\mu(\mathbf{y})$ is the total weight of the voters who hold the view $\mathbf{y}$.[12] The primary interpretation of a profile μ is as a frequency distribution of the input judgments among anoymous, hence equally weighted voters. But, more generally, μ can also incorporate weights which may reflect different voter expertise or standing.[13]

If $\boldsymbol{\mathcal{C}}=(\mathcal{K},\boldsymbol{\lambda},\mathcal{X},\mathcal{Y})$ is a judgement context and $\mu\in\Delta(\mathcal{Y})$ is a profile, then we call the pair $(\boldsymbol{\mathcal{C}};\mu)$ a *judgement problem*. A *judgement aggregation rule* on $\boldsymbol{\mathcal{C}}$ is a correspondence $F:\Delta(\mathcal{Y})\rightrightarrows\mathcal{X}$. If $F(\mu)\neq\emptyset$ for all $\mu\in\Delta(\mathcal{Y})$, then we say F is *total*. Otherwise, F is *partial*. Most judgement aggregation rules of interest are total. However, we will *not* assume that rules are total in this paper. Allowing for partiality is a simple but flexible device to accommodate restrictions on the domain of possible inputs.

In some applications, it is more appropriate to define a single judgement aggregation rule for an entire *collection* of judgement contexts, rather than a single context. For example, many rules for aggregating rankings (e.g. the Kemeny and Slater rules) or aggregating grades (e.g. Balinski and Laraki's Majority Judgement) are applicable to any number of alternatives. Likewise, many rules for selecting a committee are applicable to any number of candidates, and often have considerable flexibility with respect to the other constraints on the desired committee (e.g. minimum or maximum size). For this reason, we will often assume that the domain of a judgement aggregation rule is a (possibly infinite) *collection* $\mathfrak{C}$ of judgement contexts —we will call such a collection a *judgement environment*. (We will make this precise in Section 5.)

3.1 An example: Virtual Referendum Voting

In Virtual Referendum Voting (VRV), a unique "candidate" is to be elected as the governing agent on the basis of how well his/its position on a number of Yes-No issues with the views of the electorate. The candidate could be a person, party, or coalition of parties. The uniqueness assumption is for simplicity only.

A model of this kind has been proposed by Tangian, and is summarized in his recent monograph, *A mathematical model of democracy* (2014).[14] In the most ambitious vein, the model could be used as a blueprint for an actual voting procedure. Less ambitiously but still informatively, the model can be used as an analytical foil to compare the actual outcomes of elections to those in a hypothetical issue oriented procedure.[15] It has already received some empirical evaluation, and also ties in nicely with (typically web-based) *voter assistance systems* popular in many European countries, which are meant to inform and

[12]See formula (C1) in Appendix C for more detail.

[13]For the sake of generality, we allow profiles to be real-valued. Real-valued weights could capture anonymous aggregation with an infinite (continuum) of voters, or real-valued weights assigned to a finite set of voters. None of our results depends on this generality. In fact, a version of the main result holds also for a fixed number of anonymous voters (Theorem 3).

[14]The label "Virtual Referendum Voting" is ours, though, not his.

[15]Tangian himself is motivated by a perceived decline of representative democracy in favor of a personality oriented "audience democracy" diagnosed by influential political historian Bernard Manin (1995).

Of course, this is not the place to discuss the potential and limitations of the model in substance.

educate for voters on how the various parties align with their own views on a number of prespecified issues (Cedroni and Garzia, 2010).

In VRV, $\mathcal{K}$ is a set of political issues formulated as propositions to be approved or rejected. An element of $\{\pm 1\}^{\mathcal{K}}$ can be thought of as a *position* in ideological space. The set of issues $\mathcal{K}$ is intended to be broadly representative of the overall decision-relevant ideological space. So, in voter assistance applications, $\mathcal{K}$ typically involves between 5 and 50 issues. Voters express their views as positions, so $\mathcal{Y} = \{\pm 1\}^{\mathcal{K}}$. The candidates declare their positions as well; we will identify each candidate with her declared position (ignoring questions about the sincerity or reliability of this declaration). So $\mathcal{X}$, the available set of candidates, is typically a small some subset of $\{\pm 1\}^{\mathcal{K}}$. Some issues might be deemed more important that others, as reflected in the weight vector $\boldsymbol{\lambda}$. There is no abstract argument for issue parity here, but uniform priority weights might be sensible if the issues are selected appropriately. Clearly, the selection of the issues and their weights is itself a substantial decision. It could be made, for example, via some form of judgement aggregation by some committee of outside experts, or by the voters themselves.

The determination of the best candidate can be viewed as a problem of judgement aggregation. In particular, the best consensus candidate would be the one maintaining the "impartially most plausible" views on most positions in light of the profile of votes. The judgement aggregation rule F selecting the consensus candidates could be defined for a single context associated with a single feasible set of candidates in fixed position. But it would be more natural to let the candidates' positions vary as well, so that the judgement aggregation rule is defined for a judgement environment $\mathfrak{E}$ —say, the set of all contexts $(\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \{\pm 1\}^{\mathcal{K}})$, where $\mathcal{K}$ and $\boldsymbol{\lambda}$ are fixed and $\mathcal{X}$ is any subset of $\{\pm 1\}^{\mathcal{K}}$, or any subset of fixed cardinality m. Our main result yields an axiomatization of the general class of weighted AMRs for this model.[16]

4 Supermajority efficiency and additive majority rules

Let $(\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a judgement context. Recall that $\mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}}$ —regard $\mathcal{Y}$ as a subset of $\mathbb{R}^{\mathcal{K}}$ in the obvious way. For any profile $\mu \in \Delta(\mathcal{Y})$, we define the *majority vector* $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_k)_{k \in \mathcal{K}} \in \mathbb{R}^{\mathcal{K}}$ by setting

$$\widetilde{\mu}_k \quad := \quad \sum_{\mathbf{y} \in \mathcal{Y}} \mu(\mathbf{y}) \cdot y_k \qquad \text{for all } k \in \mathcal{K}. \tag{4.1}$$

Thus, $\widetilde{\boldsymbol{\mu}} \in \operatorname{conv}(\mathcal{Y})$ (the convex hull of $\mathcal{Y}$ in $\mathbb{R}^{\mathcal{K}}$). Now, let $q \in [0,1]$. For any $\mathbf{x} \in \mathcal{X}$, we define

$$\gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{x}}(q) \quad := \quad \sum \{\lambda_k \; ; \; k \in \mathcal{K} \text{ such that } x_k \, \widetilde{\mu}_k \geq q\}. \tag{4.2}$$

This measures the *total weight* of the issues in which the popular support for $\mathbf{x}$ exceeds the supermajority threshold q. In the special case of a uniform priority vector $\boldsymbol{\lambda} = \mathbf{1}$, we have $\gamma^{\mathbf{1}}_{\mu,\mathbf{x}}(q) = \gamma_{\mu,\mathbf{x}}(q)$, where

$$\gamma_{\mu,\mathbf{x}}(q) \quad := \quad \#\{k \in \mathcal{K} \; ; \; x_k \, \widetilde{\mu}_k \geq q\}. \tag{4.3}$$

[16] Tangian himself considered (on heuristic, not axiomatic grounds) the median rule, another AMR, and the Slater rule, which is refinable to an AMR.

This counts the *number* of issues in which the popular support for $\mathbf{x}$ exceeds the supermajority threshold q. For example, $\gamma_{\mu,\mathbf{x}}(0)$ is the number of coordinates where $\mathbf{x}$ receives at least a bare majority, $\gamma_{\mu,\mathbf{x}}(0.5)$ is the number of coordinates where $\mathbf{x}$ receives at least a 75% supermajority, and $\gamma_{\mu,\mathbf{x}}(1)$ is the number of coordinates where $\mathbf{x}$ receives unanimous support.

For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, we write "$\mathbf{x} \underline{\triangleright}^{\boldsymbol{\lambda}}_{\mu} \mathbf{y}$" ("$\mathbf{x}$ weakly supermajority-dominates $\mathbf{y}$") if $\gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{x}}(q) \geq \gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{y}}(q)$ for all $q \in (0,1]$. This relation is transitive and reflexive, but generally not complete. We write "$\mathbf{x} \stackrel{\boldsymbol{\lambda}}{\underset{\mu}{\equiv}} \mathbf{y}$" ("$\mathbf{x}$ is supermajority equivalent to $\mathbf{y}$") if $\mathbf{x} \underline{\triangleright}^{\boldsymbol{\lambda}}_{\mu} \mathbf{y}$ and $\mathbf{x} \underline{\triangleleft}^{\boldsymbol{\lambda}}_{\mu} \mathbf{y}$ (equivalently, $\gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{x}}(q) = \gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{y}}(q)$ for all $q \in (0,1]$). Finally, we write "$\mathbf{x} \triangleright^{\boldsymbol{\lambda}}_{\mu} \mathbf{y}$" ("$\mathbf{x}$ supermajority-dominates $\mathbf{y}$") if $\mathbf{x} \underline{\triangleright}^{\boldsymbol{\lambda}}_{\mu} \mathbf{y}$ but $\mathbf{x} \stackrel{\boldsymbol{\lambda}}{\underset{\mu}{\not\equiv}} \mathbf{y}$. A view $\mathbf{x}$ is *supermajority efficient* (SME) if it is maximal among the admissible output views with respect to the relation $\triangleright^{\boldsymbol{\lambda}}_{\mu}$. The set of SME views will be denoted by $\mathrm{SME}(\boldsymbol{\mathcal{C}};\mu)$. Formally,

$$\mathrm{SME}(\boldsymbol{\mathcal{C}};\mu) \quad := \quad \Big\{\mathbf{x} \in \mathcal{X}\ ;\ \mathbf{y} \triangleright^{\boldsymbol{\lambda}}_{\mu} \mathbf{x} \text{ for no } \mathbf{y} \in \mathcal{X}\Big\}.$$

In the special case of the uniform priority vector $\mathbf{1}$, we drop the "$\boldsymbol{\lambda}$" argument in these expressions, and simply write "$\mathbf{x} \triangleright_{\mu} \mathbf{y}$" etc..

Supermajority efficiency thus mandates overruling, if necessary, small and/or low-priority supermajorities in favor of an equal or larger number of supermajorities of equal or greater size and/or priority. Formally, supermajority dominance $\triangleright^{\boldsymbol{\lambda}}_{\mu}$ goes beyond Condorcet dominance by replacing an issue-wise comparison by a distributional comparison, paralleling the step from statewise dominance to first-order stochastic dominance in the theory of decision making under risk.

Example 4.1. As briefly discussed in Section 2, in the (unweighted) ranking problem with three alternatives, at any profile, all supermajority-efficient views are supermajority-equivalent, and, at any profile in which all three majority-margins differ, there is a *unique* supermajority-efficient view. Thus, at all profiles, the SME criterion can be viewed as fully decisive.

By contrast, with more than three alternatives, the SME criterion is no longer fully decisive. Formally, let $\mathcal{A} := \{a_1, \ldots, a_N\}$ be a finite set of alternatives (indexed in some arbitrary way). Let $\mathcal{K} := \{(a_n, a_m) \in \mathcal{A}^2;\ n < m\}$. Then any view $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$ can be interpreted as a complete, antisymmetric binary relation (i.e. a *tournament*) $\prec$ on $\mathcal{A}$, where $a \prec b$ if and only if either $x_{a,b} = 1$ or $x_{b,a} = -1$. (Recall that exactly one of (a,b) or (b,a) is in $\mathcal{K}$.) Now let $\mathcal{X} \subset \{\pm 1\}^{\mathcal{K}}$ be the set of views representing *transitive* tournaments (i.e. strict preference orders) on $\mathcal{A}$. The general problem of aggregating rankings over $\mathcal{A}$ can be represented as judgement aggregation on $\mathcal{X}$.

To illustrate the potential indecisiveness of SME in this context, suppose there are four alternatives, a, b, c and d, let $\mathcal{K} := \{(a,b),(a,c),(a,d),(b,c),(b,d),(c,d)\}$, and consider a profile $\mu \in \Delta(\mathcal{X})$ with majority margins $\widetilde{\mu}_{ab} = \widetilde{\mu}_{ac} = \widetilde{\mu}_{bc} = \widetilde{\mu}_{bd} = \widetilde{\mu}_{cd} = \alpha > 0$, and $\widetilde{\mu}_{ad} = \beta$.

If $\beta \geq 0$, the ranking *abcd* is the unique Condorcet consistent ranking. However, if $-\alpha \leq \beta < 0$, there are now *five* Condorcet admissible rankings (namely *abcd*, *dabc*, *cdab*, *bcda*, and *bdac*), but only one of them, namely the ranking *abcd*, is SME. (To see this, note *abcd* overrides the majority on only a single issue (namely the issue *ad*), while all other rankings override the majority on at least two issues.) Finally, if $\beta < -\alpha$, the ranking *abcd* is still SME, but so are three others, namely *dabc*, *cdab*, *bcda*. Thus, SME is not decisive in this case. $\diamondsuit$

As Example 4.1 shows, while the SME criterion is selective, it will frequently not arrive at a unique optimal view. In such cases, one must make tradeoffs between supermajorities of different sizes. A judgement aggregation rule $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ provides a systematic way of making such tradeoffs. To be precise, we will be interested in aggregation rules which satisfy the following axiom:

Axiom 1 (Supermajority efficiency) *$F(\mu) \subseteq \mathrm{SME}(\boldsymbol{\mathcal{C}}, \mu)$ for all $\mu \in \Delta(\mathcal{Y})$.*

If F satisfies Supermajority efficiency, then we will say that it is *supermajority efficient.* We will use this axiom to obtain a characterization of additive majority rules. AMRs make the tradeoff between supermajorities of various sizes in a consistent manner across profiles and contexts. How these tradeoffs are made is determined by an increasing function $\phi : [-1, +1] \longrightarrow \mathbb{R}$ (called a *gain function*). (We will later generalize this definition a bit, but the preliminary, real-valued definition, captures all the essential intuitions and calculations.) Given such a function ϕ, the *additive majority rule* $F_\phi : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is defined as follows:

$$\text{for all } \mu \in \Delta(\mathcal{Y}), \qquad F_\phi(\mu) \quad := \quad \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \left(\sum_{k \in \mathcal{K}} \lambda_k \, \phi(x_k \, \widetilde{\mu}_k) \right). \tag{4.4}$$

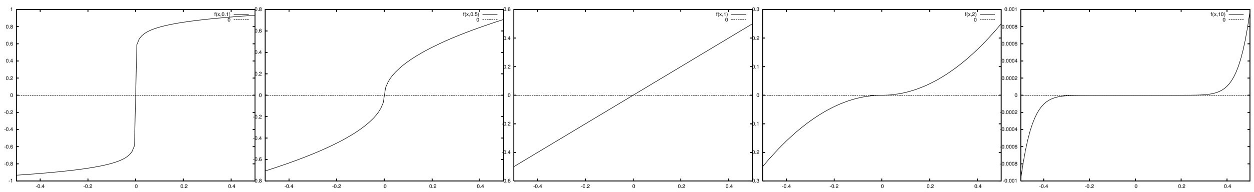
In the special case of an unweighted judgement context, this simplifies to

$$F_\phi(\mu) \quad := \quad \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \left(\sum_{k \in \mathcal{K}} \phi(x_k \, \widetilde{\mu}_k) \right). \tag{4.5}$$

Note that any additive majority rule is total. The gain function ϕ describes how the numerical strength of the majority on an issue $\widetilde{\mu}_k$ translates into its "effective strength" $\phi(\widetilde{\mu}_k)$, which determines the rate at which the majority on k is traded off against other majorities. On a broadly epistemic interpretation, $\phi(\widetilde{\mu}_k)$ is can be interpreted as reflecting the strength of the plausibility argument for aligning the consensus judgement with the majority judgement on issue k.

Qualitatively speaking, AMRs can be distinguished by the *elasticity* of the gain function: how responsive the effective strength is to the numerical strength. To illustrate, consider the following simple functional form for a gain function ϕ^d, in which $d \in (0, \infty)$ is the "degree of elasticity" of the gain function.[17] The corresponding additive majority rule $H^d := F_{\phi^d}$ is called the *homogeneous* rule of degree d. Fig. 1 illustrates the shape of the gain function for various values of d

[17] In the manner of demand theory, we can define the gain elasticity for differentiable gain-functions locally as $\eta(r) := \left| \frac{\phi'(r)}{\phi(r)} \right|$. The homogeneous gain functions are the real-valued gain functions with constant elasticity.



$\phi_{1/10}$ $\phi_{1/2}$ ϕ_1 ϕ_2 ϕ_{10}

Figure 1: Gain functions for homogeneous rules

The median rule is the unit elasticity case, with $d = 1$. At the one extreme of the range, the gain function can be highly inelastic, as illustrated by the case $d = \frac{1}{10}$. In the limit, as $d \searrow 0$, the gain function becomes essentially a step function given by $\phi(r) := \operatorname{sign}(r)$. The associated AMR F_ϕ is an adaptation of the well-known Slater rule from ranking aggregation to general judgement aggregation problems. (Note that the Slater rule itself is *not* an AMR, because $\phi = \operatorname{sign}()$ is not strictly increasing.) For such highly inelastic gain functions, the overall evaluation $\Phi(\mathbf{x}, \mu)$ of a view (as defined in formula (2.1)) primarily depends on the number of issues in which $\mathbf{x}$ is aligned with the majority judgement, while the size of majority margins matters little. By contrast, for highly elastic gain functions, small majorities matter little; the evaluation is driven primarily by the alignment of a view with the majority view on issues with a large majority.

The emphasis on large supermajorities becomes more exacerbated as d increases; in the limit, it becomes lexical; the limiting *Leximin* rule can be described as follows. For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, write "$\mathbf{x} \underset{\mu}{\approx} \mathbf{y}$" if $\gamma_{\mu,\mathbf{x}} = \gamma_{\mu,\mathbf{y}}$; otherwise write "$\mathbf{x} \underset{\mu}{\succ} \mathbf{y}$" if there exists some $Q \in (0, 1]$ such that $\gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{x}}(q) = \gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{y}}(q)$ for all $q > Q$, while $\gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{x}}(Q) > \gamma^{\boldsymbol{\lambda}}_{\mu,\mathbf{y}}(Q)$. Then $\underset{\mu}{\succeq}$ is a complete, transitive ordering of $\mathcal{X}$. We then define

$$\text{Leximin}(\boldsymbol{\mathcal{C}}, \mu) \quad := \quad \max(\mathcal{X}, \underset{\mu}{\succeq}). \tag{4.6}$$

In other words, Leximin first maximizes the total weight of coordinates which receive unanimous support (if any); then, for every possible supermajoritarian threshold $q \in (0, 1]$, Leximin maximizes the total weight number of coordinates where the support exceeds q, with higher values of q given lexicographical priority over lower ones.[18]

At first sight, the Leximin rule appears to be a natural example of an SME rule that is *not* an additive majority rule, indicating a significant limitation of this family. But this limitation is more apparent than real, since it can be overcome by allowing gain functions to be infinite and/or infinitesimally-valued —technically, by extending the co-domain of the gain function ϕ to the field $^*\mathbb{R}$ of *hyperreal numbers*.

This step may sound difficult and somewhat esoteric, but, for our purposes, the hyperreals are quite straightforward and easy to use. In a nutshell, the hyperreals are extension

[18] The Leximin rule is a refinement of the *Ranked Pairs* rule proposed by Tideman (1987) in the setting of preference aggregation; see also Zavist and Tideman (1989). The Ranked Pairs rule itself is not SME, but it agrees with the Leximin rule on profiles μ for which $\widetilde{\mu}_k \neq \widetilde{\mu}_\ell$ if $k \neq \ell$. A fortiori, the output of Leximin rule on such profiles does not depend on the priority vector $\boldsymbol{\lambda}$.

of the reals which contain infinite and infinitesimal numbers, but where nonetheless all the usual rules of arithmetic involving addition, multiplication and ordering apply.[19]

For example, using infinitesimals, one can easily define refinements of the Slater rule that are additive majority rules. Let $\psi : [-1,1] \longrightarrow \mathbb{R}$ be any strictly increasing function, let $\epsilon \in {}^*\mathbb{R}$ be an infinitesimal, and define $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ by $\phi(r) := \mathrm{sign}(r) + \epsilon\,\psi(r)$ for all $r \in [-1,1]$. Then the additive majority rule F_ϕ is a supermajority efficient refinement of the Slater rule.[20]

Thus we obtain the general, "official" definition of additive majority rules simply by changing the co-domain of the gain-function from real to hyperreal.

Definition. Let $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a judgement context. A judgement aggregation rule $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is an *additive majority rule* if there exists a strictly increasing *gain function* $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ such that, for all profiles $\mu \in \Delta(\mathcal{Y})$, we have

$$F(\mathcal{C}, \mu) \quad = \quad \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \left(\sum_{k \in \mathcal{K}} \lambda_k \, \phi(x_k \, \widetilde{\mu}_k) \right). \tag{4.7}$$

More generally, let $\mathfrak{E}$ be a judgement environment. An aggregation rule F on $\mathfrak{E}$ is an *additive majority rule* if there exists a gain function ϕ as above such that, for all $\mathcal{C} \in \mathfrak{E}$ and all $\mu \in \Delta(\mathcal{Y})$, the outcome $F(\mathcal{C}, \mu)$ is defined by formula (4.7).[21]

The next result says that we can assume without loss of generality that ϕ is odd.

Proposition 4.2 *Let $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ be any increasing gain function. Define $\widehat{\phi} : [-1,1] \longrightarrow {}^*\mathbb{R}$ by $\widehat{\phi}(r) := \phi(r) - \phi(-r)$ for $r \in [-1,1]$. Then $\widehat{\phi}$ is odd and increasing, and yields the same aggregation rule, i.e. $F_{\widehat{\phi}} = F_\phi$.*

Proposition 4.2 is due to the fact that at a given profile μ, an aggregation rule can only choose between affirming or overriding the majority of size $|\widetilde{\mu}_k|$ on a particular proposition k. Accordingly, $\widehat{\phi}(r)$ describes the gain from affirming rather than overriding a majority of the same size r on any particular proposition; it follows that the gains associated with negative majority margins are the mirror images of the gains associated with positive ones.

The following result describes some equivalent ways to represents AMRs due to oddness of the gain function. For any $\mu \in \Delta(\mathcal{Y})$, let $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_k)_{k \in \mathcal{K}}$ be defined as in formula (4.1), and then define $\phi(\widetilde{\boldsymbol{\mu}}) := (\phi(\widetilde{\mu}_k))_{k \in \mathcal{K}} \in \mathbb{R}^{\mathcal{K}}$. Then for any $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$, we define $\mathbf{x} \underset{\lambda}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) = \sum_{k \in \mathcal{K}} \lambda_k \, x_k \, \phi(\widetilde{\mu}_k)$. Let $\mathcal{M}(\mu, \mathbf{x}) := \{k \in \mathcal{K};\ \ x_k \, \widetilde{\mu}_k \geq 0\}$. This is the set of issues where $\mathbf{x}$ agrees with the majority of the voters in profile μ.

[19] Mathematically, ${}^*\mathbb{R}$ is a linearly ordered field with some additional structure that, for example, even renders exponentiation well-defined; see Appendix A for a brief introduction. No understanding of non-standard analysis is needed. Strictly speaking, there are multiple non-identical hyperreal fields, but they are all equivalent in their arithmetic rules; we construct an appropriate one in our proof.

[20] That is, $F_\phi(\mu) \subseteq \mathrm{SME}(\mathcal{C}, \mu) \cap \mathrm{Slater}(\mathcal{C}, \mu)$ for any $\mu \in \Delta(\mathcal{Y})$. *Proof:* For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $\mathbf{x} \in \mathrm{Slater}(\mathcal{C}, \mu)$ and $\mathbf{y} \notin \mathrm{Slater}(\mathcal{C}, \mu)$, then $\sum_{k \in \mathcal{K}} \mathrm{sign}(\widetilde{\mu}_k)\, x_k > \sum_{k \in \mathcal{K}} \mathrm{sign}(\widetilde{\mu}_k)\, y_k$, which implies that $\phi(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{x} > \phi(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{y}$. Thus, $\mathbf{y} \notin F_\phi(\mu)$. Thus, $F_\phi(\mu) \subseteq \mathrm{Slater}(\mathcal{C}, \mu)$. The supermajority efficiency of F_ϕ follows from Theorem 1(a) below.

[21] Here, we write "$F(\mathcal{C}, \mu)$" instead of "$F(\mu)$" to emphasize the dependence on $\mathcal{C}$ in the case when F is defined over a whole judgement environment $\mathfrak{E}$.

Proposition 4.3 *Let $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ be an odd gain function, let $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a judgement context, and let $\mu \in \Delta(\mathcal{Y})$.*

(a) $F_\phi(\mathcal{C}; \mu) = \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \sum_{k \in \mathcal{K}} \lambda_k \, \phi\,[(x_k \, \widetilde{\mu}_k)_+] = \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \sum_{k \in \mathcal{M}(\mathbf{x},\mu)} \lambda_k \, \phi \, |\widetilde{\mu}_k|$.[22]

(b) *Also,* $F_\phi(\mathcal{C}; \mu) = \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \; \mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}})$.

Additive majority rules have two basic attractive properties: they are monotone and generically single-valued. For any $\mu \in \Delta(\mathcal{Y})$, let $\mathcal{Y}(\mu) := \{\mathbf{y} \in \mathcal{Y};\; \mu(\mathbf{y}) > 0\}$. Let $\mu' \in \Delta(\mathcal{Y})$ and let $\mathbf{x} \in \mathcal{X}$. We say that μ' is ***more supportive than** μ **of*** $\mathbf{x}$ if $\mu'(\mathbf{x}) > \mu(\mathbf{x})$, while $\mu'(\mathbf{y}) < \mu(\mathbf{y})$ for all $\mathbf{y} \in \mathcal{Y}(\mu) \setminus \{\mathbf{x}\}$, and $\mu'(\mathbf{y}) = \mu(\mathbf{y}) = 0$ for all $\mathbf{y} \in \mathcal{Y} \setminus \mathcal{Y}(\mu)$. (For example: if $\delta_{\mathbf{x}} \in \Delta(\mathcal{Y})$ is the unanimous profile at $\mathbf{x}$, then for any $\mu \in \Delta(\mathcal{Y})$ and any $r \in (0,1]$, the convex combination $r\,\delta_{\mathbf{x}} + (1-r)\,\mu$ is more supportive than μ of $\mathbf{x}$.) A judgement aggregation rule $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is ***monotone*** if, for any $\mu, \mu' \in \Delta(\mathcal{Y})$, and $\mathbf{x} \in F(\mu)$, if μ' is more supportive than μ of $\mathbf{x}$, then $F(\mu') = \{\mathbf{x}\}$. In other words: if $\mathbf{x}$ is already one of the winning alternatives, then any increase in the support for $\mathbf{x}$ at the expense of support for all other elements of $\mathcal{X}$ will make $\mathbf{x}$ the *unique* winning alternative. The rule F is ***generically single-valued*** if there is an open dense subset $\mathcal{O} \subseteq \Delta(\mathcal{Y})$ such that $|F(\mu)| = 1$ for all $\mu \in \mathcal{O}$.

Proposition 4.4 *Let $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ be any gain function. For any judgement context $\mathcal{C}$, the additive majority rule F_ϕ is* (a) *monotone, and* (b) *generically single-valued on $\Delta(\mathcal{C})$.*

Technically, genericity is defined only for a continuum of individuals. Nevertheless, if the number of voters is finite but "large", then Proposition 4.4(b) can be interpreted heuristically as saying that ties are uncommon. However, if the number of voters is finite but "small", then genericity has no bearing.

5 Main result

If an aggregation rule is adopted by a group at an *ex ante*, constitutional stage, the group would often want this rule to govern multiple contexts. For example, in problems of evaluation by ranking or grading, the identity and number of alternatives (candidates) will vary. In Virtual Referendum Voting, even if the number of candidates is fixed, their position in ideological space will not be known *ex ante*. In truth-functional aggregation, the number and nature of the premises will be unknown *ex ante*, as may be the syllogism (legal doctrine) by which conclusions are derived from premises. Thus, to obtain the desired normative foundation for AMRs, we shall consider judgement aggregation rules defined on judgement environments.

The move from contexts to environments allows us to combine judgement contexts into more complex "composite" judgement contexts; this will enable us to leverage the power of the SME axiom substantially. For any contexts $\mathcal{C}_1, \ldots, \mathcal{C}_J$ with $\mathcal{C}_j = (\mathcal{K}_j, \boldsymbol{\lambda}^j, \mathcal{X}_j, \mathcal{Y}_j)$,

[22] For any $r \in \mathbb{R}$, recall that $r_+ := \max\{r, 0\}$.

their combination ("product") $\mathcal{C}_1 \times \cdots \times \mathcal{C}_J$ is given by the context $(\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ where $\mathcal{K} := \mathcal{K}_1 \sqcup \cdots \sqcup \mathcal{K}_J$, $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_J$, $\mathcal{Y} := \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_J$ and $\boldsymbol{\lambda} := (\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^J)$ (regarded as an element of $\mathbb{R}_+^{\mathcal{K}} \cong \mathbb{R}_+^{\mathcal{K}_1} \times \cdots \times \mathbb{R}_+^{\mathcal{K}_J}$). Thus, composite contexts are obtained by taking the Cartesian products of input spaces and the output spaces; the weights of any issue in the composite context are simply given by the weight of the issue in the component context. In particular, if each component context has unit weight vector $\mathbf{1} \in \mathbb{R}^{\mathcal{K}_j}$, then the composite context also has unit weight vector $\mathbf{1} \in \mathbb{R}^{\mathcal{K}}$, the same scale of each component context reflecting "equal importance" of each. In this manner, our definition of products in the general, weighted case also yields a self-contained notion of product for the unweighted (i.e. unit-weighted) case.

The combination of contexts applies to situations in which a group is faced with several judgement problems jointly. For example, a hiring committee might have to make decisions on different positions $j \in [1 \ldots J]$ with disjoint sets of applicants $\mathcal{A}_j$. For each position, the committee members would submit a ranking of applicants on that position. The component problem being characterized by issues $\mathcal{K}_j = \mathcal{A}_j \times \mathcal{A}_j$, the composite problem has the issue space $\mathcal{K} := \bigsqcup_{j=1}^J \mathcal{K}_j$, and the admissible input as well as output judgements are given as J-tuples of rankings (rather than as one overall ranking of all the candidates, which would result in issue space $(\bigsqcup_{j=1}^J \mathcal{A}_j) \times (\bigsqcup_{j=1}^J \mathcal{A}_j)$).

We will require a judgement rule to pronounce on the best group judgement(s) not just for the judgement contexts in the environment $\mathfrak{E}$, but for all contexts that arise from the arbitrary *combinations* of contexts in $\mathfrak{E}$. We shall refer to this set as the ***closure*** (under combination) of $\mathfrak{E}$, and denote it by $\langle \mathfrak{E} \rangle$. Formally,

$$\langle \mathfrak{E} \rangle \quad := \quad \{\mathcal{C}_1 \times \cdots \times \mathcal{C}_J \; ; \; J \in \mathbb{N} \text{ and } \; \mathcal{C}_1, \ldots, \mathcal{C}_J \in \mathfrak{E}\}. \tag{5.1}$$

Evidently, the environment $\langle \mathfrak{E} \rangle$ is itself closed under the combination of contexts, and is the smallest collection with this property. For any judgement context $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$, let $o(\mathcal{C}) := \mathcal{X}$ and $\Delta(\mathcal{C}) := \Delta(\mathcal{Y})$. For any judgement environment $\mathfrak{E}$, with closure $\langle \mathfrak{E} \rangle$, let

$$\Delta\langle \mathfrak{E} \rangle \quad := \quad \Big\{(\mathcal{C}; \mu) \; ; \; \mathcal{C} \in \langle \mathfrak{E} \rangle \text{ and } \; \mu \in \Delta(\mathcal{C})\Big\}, \quad \text{and} \quad o\langle \mathfrak{E} \rangle \quad := \bigcup_{\mathcal{C} \in \langle \mathfrak{E} \rangle} o(\mathcal{C}). \tag{5.2}$$

A ***judgement aggregation rule*** on $\langle \mathfrak{E} \rangle$ is a correspondence $F : \Delta\langle \mathfrak{E} \rangle \rightrightarrows o\langle \mathfrak{E} \rangle$ such that, for any $\mathcal{C} \in \langle \mathfrak{E} \rangle$, we have $F(\mathcal{C}, \mu) \subseteq o(\mathcal{C})$ for all $\mu \in \Delta(\mathcal{C})$.

If $\mathcal{Y} := \mathcal{Y}_1 \times ... \times \mathcal{Y}_J$, and $\mu \in \Delta(\mathcal{Y})$, then for all $j \in [1 \ldots J]$, let $\mu^{(j)} \in \Delta(\mathcal{Y}_j)$ be the marginal projection of μ onto $\mathcal{Y}_j$. Formally,

$$\text{for all } \mathbf{x} \in \mathcal{Y}_j, \qquad \mu^{(j)}(\mathbf{x}) \quad := \quad \sum_{\substack{(\mathbf{y}_1, \ldots, \mathbf{y}_J) \in \mathcal{Y} \\ \text{with } \mathbf{y}_j = \mathbf{x}}} \mu(\mathbf{y}_1, \ldots, \mathbf{y}_J). \tag{5.3}$$

Observe that $\widetilde{\boldsymbol{\mu}} = (\widetilde{\boldsymbol{\mu}}^{(1)}, \ldots, \widetilde{\boldsymbol{\mu}}^{(J)})$.

We require that the group judgement on composite problems is given by combining the optimal group judgements on each constituent problem. This is expressed by the following axiom.

Axiom 2 (Combination) *For all contexts* $\mathcal{C}_1, ..., \mathcal{C}_J \in \mathfrak{E}$ *and* $\mathcal{C} = \mathcal{C}_1 \times ... \times \mathcal{C}_J$ *and all* $\mu \in \Delta(\mathcal{C})$, $F(\mathcal{C}, \mu) = F(\mathcal{C}_1, \mu^{(1)}) \times ... \times F(\mathcal{C}_J, \mu^{(J)})$.

From the broadly epistemic perspective aiming at consensus, the Combination axiom is quite compelling, independently of whether consensus is to be construed in a majoritarian manner or not. The group is faced simultaneously with a number of logically independent judgement tasks; the relevant evidence for each component judgement $\mathbf{x}_j$ is given by voters' views on that component judgement. Hence the the "most plausible" group judgement yielding the best supported consensus view can be determined by simply combining the componentwise most plausible judgements independently.[23]

For the Combination axiom to be normatively relevant, it suffices for the combination of judgement aggregation problems to refer to well-defined hypotheticals; whether or not combined judgement problems are of immediate interest is of no particular relevance. At the same time, combinations of judgement problems arise naturally in actuality within temporal settings, when the group faces a sequence of judgement problems over time. The Combination axiom then says the entire equence of group judgements at each date must be optimal – according to F – when considered jointly. In particular, the combined judgement must satisfy any normative restrictions applying to the composite context $(\mathcal{C}, \mu)$ —in particular, Supermajority Efficiency. This multiplies the power of the SME axiom; indeed, these two axioms alone almost suffice to ensure representation as an AMR.

We have now all the ingredients for the main result for environments with rational-valued weights, hence in particular for unweighted environments. For real-valued weights, one must strengthen the requirements on the environment and the aggregation rule a bit. Specifically, we must assume that it is possible to at least slightly vary the weight of each context relative to other contexts in $\mathfrak{E}$. Formally, for any judgement context $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ and any $r > 0$, we define $r\,\mathcal{C} := (\mathcal{K}, r\,\boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. We will assume that the environment $\mathfrak{E}$ is *minimally rich*, i.e. that, for any $\mathcal{C} \in \mathfrak{E}$, there exists ϵ such that, $r\,\mathcal{C} \in \mathfrak{E}$ for all $r \in (1-\epsilon, 1+\epsilon)$. However, such a change in scale should not affect the output of the rule *within* a particular context, since it leaves the relative weights of issues within that context unchanged. This is the content of the next axiom.

Axiom 3 (Scale invariance) *Let* $\mathcal{C} \in \mathfrak{E}$, *let* $r > 0$, *and suppose* $r\,\mathcal{C} \in \mathfrak{E}$ *also. Then for any* $\mu \in \Delta(\mathcal{C})$, *we have* $F(\mathcal{C}, \mu) = F(r\,\mathcal{C}, \mu)$.

Note that AMRs are obviously scale-invariant, as is the SME set itself. We say that the rule G *covers* the rule F if, for any $\mu \in \Delta(\mathcal{C})$, $F(\mathcal{C}, \mu) \subseteq G(\mathcal{C}, \mu)$. The rule G *minimally covers* F if G is covered by any other rule that covers F. The following is our main result.

Theorem 1 *Let* $\mathfrak{E}$ *be a judgement environment.*

[23]The Combination axiom is also sensible from a compromise perspective, but perhaps more open to qualification. It makes sense to posit, as done by the axiom, that the aggregate burden of compromise is minimized iff it is minimized component-by-component. The Combination axiom rules out, however, distributional considerations which would take into account whether it is the same individuals bearing much of the "burden of compromise" in each instance, or whether that burden is distributed more equally over component problems.

(a) *Any additive majority rule satisfies* Supermajority efficiency, Combination *and* Scale invariance *on* $\langle\mathfrak{E}\rangle$.

Now let F be a judgement aggregation rule satisfying Combination *on* $\langle\mathfrak{E}\rangle$. *Also suppose that either* (i) $\mathfrak{E}$ *is a rationally-weighted environment, or* (ii) $\mathfrak{E}$ *is minimally rich, and F satisfies* Scale invariance. *Then:*

(b) *F satisfies* Supermajority efficiency *on $\Delta\langle\mathfrak{E}\rangle$ if and only if F is covered by an additive majority rule G.*

Now, in addition to the previous assumptions, suppose F is SME and total on $\langle\mathfrak{E}\rangle$. Then:

(c) *For all $\boldsymbol{\mathcal{C}} \in \langle\mathfrak{E}\rangle$, there is a dense open subset $\mathcal{O} \subseteq \Delta(\boldsymbol{\mathcal{C}})$ such that $F(\boldsymbol{\mathcal{C}}, \mu) = G(\boldsymbol{\mathcal{C}}; \mu)$ and is single-valued for all $\mu \in \mathcal{O}$.*

(d) *F admits a unique minimal cover in the class of additive majority rules.*

Theorem 1 is a multi-faceted statement. To distill its gist, consider the special case of the environment $\mathfrak{E} = \langle\{\boldsymbol{\mathcal{C}}\}\rangle$ generated by a single rationally weighted judgement context $\boldsymbol{\mathcal{C}}$. If F is any judgement aggregation rule on $\boldsymbol{\mathcal{C}}$, then there exists a unique extension of F to a judgement aggregation rule F^* on the environment $\langle\{\boldsymbol{\mathcal{C}}\}\rangle$ that satisfies Combination. This yields the following consequence of Theorem 1.

Corollary 5.1 *Let F be a judgement aggregation rule on a rationally-weighted context $\boldsymbol{\mathcal{C}}$. Then F is covered by an additive majority rule if and only if F^* satisfies* Supermajority efficiency *on* $\langle\{\boldsymbol{\mathcal{C}}\}\rangle$.

The Combination axiom is an interprofile condition. As such, it is very transparent, even plain, and compatible with many different JA rules. At the same time, as Theorem 1 and Corollary 5.1 show, it is quite potent in conjunction with Supermajority Efficiency. This is possible because application of Supermajority Efficiency to composite judgement contexts entails substantial additional restrictions beyond the application of Supermajority Efficiency to the constituent judgement contexts.

To illustrate how this happens, consider the Monotonicity condition satisfied by all AMRs, as shown by Proposition 4.4. Monotonicity is a basic, intrinsically appealing, cross-profile restriction on judgement aggregation rules that is evidently not entailed by Supermajority Efficiency alone. Arguing by way of contradiction, consider any context $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ and profiles $\mu, \mu' \in \Delta(\mathcal{Y})$ witnessing a violation of monotonicity. That is, assume that while μ' is more supportive of some $\mathbf{x}$ than μ is (as defined above in Section 4), $\mathbf{x} \in F(\mu)$ but $\mathbf{x} \notin F(\mu')$, and say $\mathbf{y} \in F(\mu')$. Consider the combined context $\widehat{\boldsymbol{\mathcal{C}}} = \boldsymbol{\mathcal{C}} \times \boldsymbol{\mathcal{C}}$ and a profile $\widehat{\mu}$ with marginals $\widehat{\mu}^{(1)} = \mu$ and $\widehat{\mu}^{(2)} = \mu'$, respectively. By Combination, $(\mathbf{x}, \mathbf{y}) \in F(\widehat{\boldsymbol{\mathcal{C}}}, \widehat{\mu})$. But $(\mathbf{x}, \mathbf{y})$ is supermajority-dominated in $(\widehat{\boldsymbol{\mathcal{C}}}, \widehat{\mu})$ by $(\mathbf{y}, \mathbf{x})$, as follows easily from the fact that $(\mu' - \mu)_k \cdot (x_k - y_k) > 0$ for all $k \in \mathcal{K}$ such that $x_k \neq y_k$, by hypothesis. So F must be monotone after all.

The Monotonicity condition illustrates the bite of Combination for a single context. By applying Supermajority Efficiency and Combination across different contexts, we also ensure that the supermajority tradeoffs are are made in a uniform way across all contexts, being governed by the same gain function ϕ throughout.[24]

5.1 Continuity

Theorem 1 is slightly unsatisfactory as its representation by AMRs is an inclusion rather than an equality. That gap can be closed if F is assumed to be continuous and total. A judgement aggregation rule $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is *continuous* if, for every $\mu \in \Delta(\mathcal{Y})$, one of the following two equivalent statements is true:

(Cont1) There exists some $\overline{\epsilon} > 0$ such that, for any $\epsilon \in (0, \overline{\epsilon})$ any other $\nu \in \Delta(\mathcal{Y})$, we have $F(\epsilon\, \nu + (1-\epsilon)\, \mu) \subseteq F(\mu)$.

(Cont2) For every sequence $\{\mu_n\}_{n=1}^{\infty} \subset \Delta(\mathcal{Y})$, and every $\mathbf{x} \in \mathcal{X}$, if $\lim_{n\to\infty} \mu_n = \mu$, and $\mathbf{x} \in F(\mu_n)$ for all $n \in \mathbb{N}$, then $\mathbf{x} \in F(\mu)$.

Statement (Cont1) is sometimes described as the "overwhelming majority" property. Heuristically, the profile $\epsilon\, \nu + (1-\epsilon)\, \mu$ represents a mixture of two populations: a small minority described by the profile ν, and a large majority represented by the profile μ. Statement (Cont1) says that, if the majority is large enough, then its view determines the group view (except that the minority can perhaps act as a tie-breaker in some cases). The equivalent statement (Cont2) is a standard *upper hemicontinuity* property.

Theorem 2 *Let F be a total and continuous aggregation rule on an environment $\langle \mathfrak{E} \rangle$ satisfying either condition* (i) *or condition* (ii) *of Theorem 1. Then F is an additive majority rule if and only if F satisfies* Supermajority efficiency *and* Combination.

Let $\mathcal{C}$ be a judgement context, and let F be a judgement aggregation rule on $\mathcal{C}$. If $\mathcal{C}$ is rationally weighted, then we define $\langle \mathcal{C} \rangle := \langle \{\mathcal{C}\} \rangle$; in this cases, as we already noted prior to Corollary 5.1, there is a unique extension of F to a judgement aggregation rule F^* on $\langle \mathcal{C} \rangle$ that satisfies Combination. On the other hand, if $\mathcal{C}$ is *not* rationally weighted, then we define $\langle \mathcal{C} \rangle := \langle \{r\, \mathcal{C};\ r \in \mathbb{R}_+\} \rangle$; in this cases, there is a unique extension of F to a rule F^* on $\langle \mathcal{C} \rangle$ that satisfies both Combination and Scale invariance. We can now state the following enhancement of Corollary 5.1.

Corollary 5.2 *Let F be a continuous judgement aggregation rule on $\mathcal{C}$. Then F^* satisfies* Supermajority efficiency *on $\langle \mathcal{C} \rangle$ if and only if F is an additive majority rule.*

While Theorem 2 delivers an exact representation of continuous JA rules as AMRs, it remain silent on *when* an AMR is continuous. A straightforward sufficient condition is the real-valuedness and continuity of the gain function.

---

[24] We have stated the axiom of Combination in terms of arbitrary profiles in $\Delta(\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_J)$. But to obtain the conclusions of Theorem 1, we only need Combination to hold for profiles of the form $\mu_1 \times \cdots \times \mu_J$, where $\mu_j \in \Delta(\mathcal{Y}_j)$ for all $j \in [1 \ldots J]$. (This sort of "Cartesian product" could arise in a profile involving J disjoint populations of voters, for example.) We opted for the stronger forms of Combination, because it remains applicable in the finite population case as well.

Proposition 5.3 *Let $\phi : [-1, 1] \longrightarrow \mathbb{R}$ be any continuous, real-valued gain function. Then for every judgement context $\boldsymbol{\mathcal{C}}$, the rule F_ϕ is continuous on $\Delta(\boldsymbol{\mathcal{C}})$.*

Following Myerson (1995), one might hope to obtain a converse to Proposition 5.3, and thereby to turn Theorem 1 into a full characterization of continuous AMR. But one can show by example that the converse is false. Indeed, continuity of an AMR does not even ensure representability by a real-valued gain function. So, the characterization of AMRs that are continuous, of AMRs that have a real-valued representation, and of AMRs that have a continuous, real-valued representation are all open problems. The difficulty in obtaining a real-valued representation also explains why it was necessary to go the hyperreal route rather than try to obtain a real-valued representation via a more standard argument from infinite-dimensional analysis.[25]

5.2 Restricted Domains; Finite Populations

The central assertion of Theorem 1 is part (b): the characterization of SME in terms of the existence of a covering AMR. Note that this does not require the rule to be total, i.e. non-empty everywhere. (Indeed, the assertion is trivially true of the "null" correspondence that selects the empty set everywhere.) This feature lends additional versatility to result.

For example, Theorem 1(b) holds also when the assumption that $\mathcal{X} \subseteq \mathcal{Y}$ is dropped. This can capture restrictions on admissible input judgements that are *not* imposed on output judgements. For instance, in the context of aggregating rankings, the input rankings may have certain structure (e.g. for reasons of informational parsimony). At the same time, one might be satisfied with any (transitive and complete) ranking as an admissible output. The insensitivity of our main result to the scope of the domain of is quite unusual in social choice theory.[26]

In similar manner, Theorem 1 also applies to fixed finite populations of size N, as shown in the following result. Assuming anonymity, for a given input space $\mathcal{Y}$, let $\Delta_N(\mathcal{Y})$ denote the possible judgement profiles among N equally weighted voters,[27] and define $\Delta_N(\langle\mathfrak{E}\rangle)$ analogously.

Theorem 3 *Let $\mathfrak{E}$ be a finite, rationally weighted judgement environment, and let N be a positive integer. Let F be an aggregation rule on the finite population domain $\Delta_N(\langle\mathfrak{E}\rangle)$ satisfying* Combination. *Then F satisfies* Supermajority efficiency *if and only if it is covered by an additive majority rule G. If F is total, it admits a unique minimal cover. Finally, any additive majority rule on $\Delta_N(\langle\mathfrak{E}\rangle)$ can be represented by an additive majority rule with a real-valued gain function.*

[25] With an infinite-dimensional separating hyerplane argument analogous to the one used in the proofs of Theorems C.1 and C.2 in Appendix C, we *can* construct a function $\phi : [-1, 1] \longrightarrow \mathbb{R}$ such that F_ϕ covers F. However, in general, this ϕ will *not* be strictly increasing, so F_ϕ will not, in general, be SME itself. Indeed, in extreme cases (e.g. F = Leximin), ϕ may be constant on almost all of $[-1, 1]$, so this cover will be so coarse as to be useless. Note that this observation also underlines the key role of *strict* monotonicity of ϕ for generic single-valuedness shown in Proposition 4.4.

[26] It contrasts, for example, with the reliance on large, if not unrestricted, domains in results such as Arrow's Impossibility Theorem and the characterization of positonal scoring rules due to Young (1975).

[27] Formally, $\Delta_N(\mathcal{Y}) = \{\mu \in \Delta(\mathcal{Y});\ \text{for all } \mathbf{y} \in \mathcal{Y},\ \mu(\mathbf{y}) = m/N \text{ for some } m \in [0 \ldots M]\}$.

The proof of our main result (Theorem 1) begins by proving a version of Theorem 3 for a fixed, finite population of weighted voters (Theorem C.1). We also obtain a version of Theorem 3 for *infinite*, *non*-rationally weighted judgement environments (Theorem C.2), but it is more complicated to state. We then patch together the real-valued representations obtained via Theorems C.1 and C.2 for every possible finitely-based subdomain, to obtain one overarching representation with a common gain function. This common gain function is hyperreal-valued because it is obtained from an ultrapower construction; see Appendix A for details.

Indeed, the need for hyperreal numbers already arises if one wants to deal with the variable-population domain $\bigcup_{N\in\mathbb{N}} \Delta_N(\mathfrak{E})$, which is the domain of all *rational*-valued profiles. For example, consider the Leximin rule. By the concluding statement of Theorem 3, for any $N \in \mathbb{N}$, Leximin is covered on $\Delta_N(\langle\mathfrak{E}\rangle)$ by an AMR with a real-valued gain function. Indeed, it is easy to check that Leximin $= H^{d_N}$ on $\Delta_N(\mathfrak{E})$ for d_N sufficiently large. But as $N \to \infty$, this d_N must increase without bounds. Hence, there is no $d \in \mathbb{R}$ such that Leximin $= H^d$ on $\bigcup_{N\in\mathbb{N}} \Delta_N(\mathfrak{E})$. Indeed, there is *no* real-valued gain function ϕ such that Leximin $= F_\phi$ on $\Delta_N(\mathfrak{E})$ for all $N \in \mathbb{N}$. However, if ω is an *infinite* hyperreal number, then H^ω is well-defined and equal to Leximin. (For the proofs of these statements, see Proposition F.1 in Appendix F.)

Conclusion

In this paper, we have characterized majoritarian judgement aggregation in terms of the family of additive majority rules. We have provided a representation theorem with two types of free parameters: the issue weights and the gain function. But we have not said much about how these free parameters should be determined.

As to issue weights, symmetries of the judgment context are often suggestive of symmetry of the weights; can one provide a formally explicit account of when such arguments are cogent? Also, we have assumed that issue weights are given. When is this assumption sensible? In special cases, one may attempt to obtain the issue weights themselves from a representation result, but in general, that seems to be too much to ask for.

Likewise, are there further normative considerations that could pin down a particular gain function, or a subset of them? We believe that there are such considerations, but that they do not all pull in one direction; substantial work remains to be done. A central place in these investigations will surely be played by the rule associated with any linear gain function —namely the median rule. A normative, axiomatic foundation of this rule is provided in the companion paper (Nehring and Pivato, 2018).

Appendices

A Hyperreal fields

In basic mathematics, infinity is a "nonarithmetic" object, and the word "infinitesimal" is merely a figure of speech. But it is possible to construct a well-defined and well-behaved arithmetic of infinite and infinitesimal quantities, using a *hyperreal field*. Roughly speaking, this is is an arithmetic structure ${}^*\mathbb{R}$ which is obtained by adding "infinite" and "infinitesimal" quantities to the set of real numbers. The important properties of ${}^*\mathbb{R}$ are as follows:

1. ${}^*\mathbb{R}$ is a *field*. This means that ${}^*\mathbb{R}$ has binary operations "+" and "$\cdot$", and distinguished "identity" elements 0 and 1 such that:

   (a) $({}^*\mathbb{R}, +, 0)$ is an abelian group, and $({}^*\mathbb{R} \setminus \{0\}, \cdot, 1)$ is an abelian group.

   (b) For all $r, s, t \in {}^*\mathbb{R}$, we have $r \cdot (s + t) = r \cdot s + r \cdot t$.

2. There is an "exponentiation" operation, which behaves as one would expect: for all $r, s, t \in {}^*\mathbb{R}$, we have $r^0 = 1$, $r^1 = r$, $r^{-1} = 1/r$, $r^{s+t} = r^s \cdot r^t$, and $(r \cdot s)^t = r^t \cdot s^t$.

3. ${}^*\mathbb{R}$ has a linear order relation $>$ (i.e. $>$ is transitive, complete, and antisymmetric).

4. For any $r, s \in {}^*\mathbb{R}$, if $r > 0$, then $r + s > s$. If $r > 1$, and $s > 0$, then $r \cdot s > s$.

5. $\mathbb{R} \subset {}^*\mathbb{R}$. The arithmetic and ordering on ${}^*\mathbb{R}$ extend the arithmetic and ordering of $\mathbb{R}$.

The field ${}^*\mathbb{R}$ is the subject of of *nonstandard analysis* (Goldblatt, 1998). It inherits many of the properties of $\mathbb{R}$, but not all. For example, because it contains infinite quantities, ${}^*\mathbb{R}$ violates the *Archimedean* property of the real numbers, which states that the ratio of any two non-zero elements is finite. But this non-Archimidean aspect is crucial to allow representation of AMRs with lexical features such as the Leximin rule and SME refinements of the Slater rule. Also, because ${}^*\mathbb{R}$ contains infinitesimals, it is not *order-complete*: in general, infinite subsets of ${}^*\mathbb{R}$ need not have well-defined suprema and infima. Thus, much of the machinery of classical real analysis breaks down in ${}^*\mathbb{R}$. The order topology of ${}^*\mathbb{R}$ is not well-behaved; there are no nontrivial continuous functions from $\mathbb{R}$ into ${}^*\mathbb{R}$. While these features create difficulties in asserting and proving some statements, they do not affect the use of hyperreal numbers in the calculus of additive majority rules since the sets involved there are always finite (at least as long as the number of issues is finite, as assumed here.)

So our use of hyperreals stays elementary throughout. The only "non-elementary" argument is the ultra-power construction which appeals to the Axiom of Choice. No knowledge of nonstandard analysis is required.

Ultrapower construction. The properties listed above are sufficient for a casual user of ${}^*\mathbb{R}$. However, we will now also provide a formal construction of ${}^*\mathbb{R}$, that is central to the proof of the main result of the paper, Theorem 1. Let $\mathcal{I}$ be any infinite indexing set. A *free filter* on $\mathcal{I}$ is a collection $\mathfrak{F}$ of subsets of $\mathcal{I}$ satisfying the following axioms:

(F0) No finite subset of $\mathcal{I}$ is an element of $\mathfrak{F}$. (In particular, $\emptyset \notin \mathfrak{F}$.)

(F1) If $\mathcal{E}, \mathcal{F} \in \mathfrak{F}$, then $\mathcal{E} \cap \mathcal{F} \in \mathfrak{F}$.

(F2) For any $\mathcal{F} \in \mathfrak{F}$ and $\mathcal{E} \subseteq \mathcal{I}$, if $\mathcal{F} \subseteq \mathcal{E}$, then $\mathcal{E} \in \mathfrak{F}$ also.

For any $\mathcal{E} \subseteq \mathcal{I}$, axioms (F0) and (F1) together imply that at most one of $\mathcal{E}$ or $\mathcal{E}^{\complement} := \mathcal{I} \setminus \mathcal{E}$ can be in $\mathfrak{F}$. A *free ultrafilter* is filter $\mathfrak{U}$ which also satisfies:

(UF) For any $\mathcal{E} \subseteq \mathcal{I}$, either $\mathcal{E} \in \mathfrak{U}$ or $\mathcal{E}^{\complement} \in \mathfrak{U}$.

Equivalently, a free ultrafilter is a maximal free filter: it is not a proper subset of any other free filter. Heuristically, elements of $\mathfrak{U}$ are "large" subsets of $\mathcal{I}$: if $\mathcal{J} \in \mathfrak{U}$ and a certain statement holds for all $j \in \mathcal{J}$, then this statement holds for "almost all" element of $\mathcal{I}$. In particular, axioms (F0) and (UF) imply that $\mathcal{I} \in \mathfrak{U}$.

Ultrafilter lemma. *Every free filter $\mathfrak{F}$ is contained in some free ultrafilter.*

Proof sketch. Consider the set of all free filters containing $\mathfrak{F}$; apply Zorn's Lemma to get a maximal element of this set. $\square$

Let $\mathfrak{U}$ be a free ultrafilter on $\mathcal{I}$, and let $\mathbb{R}^{\mathcal{I}}$ be the space of all functions $r : \mathcal{I} \longrightarrow \mathbb{R}$. For all $r, s \in \mathbb{R}^{\mathcal{I}}$, define $r \underset{\mathfrak{U}}{\approx} s$ if the set $\{i \in \mathcal{I};\, r(i) = s(i)\}$ is an element of $\mathfrak{U}$. Let ${}^*\mathbb{R} := \mathbb{R}^{\mathcal{I}}/(\underset{\mathfrak{U}}{\approx})$. For any $r \in \mathbb{R}^{\mathcal{I}}$, let ${}^*r$ denote the equivalence class of r in ${}^*\mathbb{R}$. For any $r, s \in \mathbb{R}^{\mathcal{I}}$, we set

$$\left({}^*r > {}^*s\right) \quad \Longleftrightarrow \quad \left(\text{the set } \{i \in \mathcal{I};\, r(i) > s(i)\} \text{ is an element of } \mathfrak{U}\right). \tag{A1}$$

This defines a linear order ">" on ${}^*\mathbb{R}$. Define the elements $r + s$, $r \cdot s$, r/s, and r^s in $\mathbb{R}^{\mathcal{I}}$ by: $(r+s)_i := r_i + s_i$, $(r \cdot s)_i := r_i \cdot s_i$, $(r/s)_i := r_i/s_i$, and $(r^s)_i := r_i^{s_i}$, for all $i \in \mathcal{I}$. Then, define ${}^*r + {}^*s := {}^*(r+s)$, ${}^*r \cdot {}^*s := {}^*(r \cdot s)$, ${}^*r/{}^*s := {}^*(r/s)$, and ${}^*r^{{}^*s} := {}^*(r^s)$. We embed $\mathbb{R}$ into ${}^*\mathbb{R}$ by mapping each $r \in \mathbb{R}$ to the element ${}^*\overline{r}$ in ${}^*\mathbb{R}$, where $\overline{r} := (r, r, r, \ldots) \in \mathbb{R}^{\mathcal{I}}$. Then $({}^*\mathbb{R}, +, \cdot, >)$ is called an *ultrapower* of $\mathbb{R}$; it is a hyperreal field in the sense defined above.

B Proofs from Section 4

Notation. For any $\mathbf{x}, \mathbf{y} \in \{\pm 1\}^{\mathcal{K}}$, we define $\mathcal{K}(\mathbf{x}, \mathbf{y}) := \{k \in \mathcal{K}\,;\ x_k \neq y_k\}$.

Proof of Proposition 4.2. (a) Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. Define $\widehat{\phi}(r) := \phi(r) - \phi(-r)$ for all $r \in [-1, 1]$. Then $\widehat{\phi}$ is odd. For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, and any $\mu \in \Delta(\mathcal{Y})$, we have

$$\begin{aligned}
\sum_{k \in \mathcal{K}} \lambda_k\, \phi(x_k \widetilde{\mu}_k) - \sum_{k \in \mathcal{K}} \lambda_k\, \phi(y_k \widetilde{\mu}_k) \quad &= \quad \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, \Big(\phi(x_k \widetilde{\mu}_k) - \phi(y_k \widetilde{\mu}_k)\Big) \\
\underset{(*)}{=} \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, \Big(\phi(x_k \widetilde{\mu}_k) - \phi(-x_k \widetilde{\mu}_k)\Big) \quad &\underset{(\dagger)}{=} \quad \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, \widehat{\phi}(x_k \widetilde{\mu}_k) \\
\underset{(\ddagger)}{=} \frac{1}{2} \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, \Big(\widehat{\phi}(x_k \widetilde{\mu}_k) - \widehat{\phi}\,(-x_k \widetilde{\mu}_k)\Big) \quad &\underset{(*)}{=} \quad \frac{1}{2} \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, \Big(\widehat{\phi}(x_k \widetilde{\mu}_k) - \widehat{\phi}(y_k \widetilde{\mu}_k)\Big) \\
= \frac{1}{2} \left(\sum_{k \in \mathcal{K}} \lambda_k\, \widehat{\phi}(x_k \widetilde{\mu}_k) - \sum_{k \in \mathcal{K}} \lambda_k\, \widehat{\phi}(y_k \widetilde{\mu}_k)\right). &
\end{aligned} \tag{B1}$$

Here $(*)$ is because $y_k = -x_k$ for all $k \in \mathcal{K}(\mathbf{x}, \mathbf{y})$. Meanwhile $(\dagger)$ is simply the definition of $\widehat{\phi}$, and $(\ddagger)$ is because $\widehat{\phi}$ is odd, so $\widehat{\phi}(r) - \widehat{\phi}(-r) = 2\widehat{\phi}(r)$.

Equation (B1) implies that

$$\left(\sum_{k\in\mathcal{K}} \lambda_k\, \phi(x_k\widetilde{\mu}_k) \geq \sum_{k\in\mathcal{K}} \lambda_k\, \phi(y_k\widetilde{\mu}_k)\right) \iff \left(\sum_{k\in\mathcal{K}} \lambda_k\, \widehat{\phi}(x_k\widetilde{\mu}_k) \geq \sum_{k\in\mathcal{K}} \lambda_k\, \widehat{\phi}(y_k\widetilde{\mu}_k)\right).$$

This holds for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Thus, $F_\phi(\boldsymbol{\mathcal{C}}; \mu) = F_{\widehat{\phi}}(\boldsymbol{\mathcal{C}}; \mu)$. □

Proof of Proposition 4.3. (a) The second equality follows immediately from the first, because $x_k\widetilde{\mu}_k > 0$ if and only if $k \in \mathcal{M}(\mathbf{x}, \mu)$. To see the first equality, define $\psi(r) := \phi(r_+)$ for all $r \in [-1, 1]$. Then $\phi(r) = \psi(r) - \psi(-r)$ for all $r \in [-1, 1]$ (because ϕ is odd). Thus, in the notation of Proposition 4.2, we have $\phi = \widehat{\psi}$. Thus, $F_\phi = F_\psi$. (Strictly speaking, ψ is only non-decreasing, rather than strictly increasing. But the proof of Proposition 4.2 still works.)

(b) If ϕ is odd, then for any $\mu \in \Delta(\mathcal{Y})$, $\mathbf{x} \in \mathcal{X}$, and $k \in \mathcal{K}$, we have $\phi(x_k\widetilde{\mu}_k) = x_k\phi(\widetilde{\mu}_k)$ (because $x_k = \pm 1$). Thus, $\displaystyle\sum_{k\in\mathcal{K}} \lambda_k\, \phi(x_k \cdot \widetilde{\mu}_k) = \sum_{k\in\mathcal{K}} \lambda_k\, x_k\, \phi(\widetilde{\mu}_k) = \mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}})$. Thus,

$$\operatorname*{argmax}_{\mathbf{x}\in\mathcal{X}} \left(\sum_{k\in\mathcal{K}} \lambda_k\, \phi(x_k\, \widetilde{\mu}_k)\right) = \operatorname*{argmax}_{\mathbf{x}\in\mathcal{X}}\ \mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}).$$ □

Lemma B.1 *Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a judgement context, and let $F : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be an aggregation rule. If F is monotone and total, then F is generically single-valued on $\Delta(\mathcal{Y})$.*

Proof: Define $\Delta^*(\mathcal{Y}) := \{\mu \in \Delta(\mathcal{Y});\ \ \mu(\mathbf{y}) > 0$ for all $\mathbf{y} \in \mathcal{Y}\}$ (a dense subset of $\Delta(\mathcal{Y})$). For any $\mu \in \Delta^*(\mathcal{Y})$ and any $\mathbf{x} \in F(\mu)$, define $\mathcal{O}_\mathbf{x}(\mu) := \{\mu' \in \Delta(\mathcal{Y});\ \ \mu'(\mathbf{x}) > \mu(\mathbf{x})$ while $\mu'(\mathbf{y}) < \mu(\mathbf{y})$, for all $\mathbf{y} \in \mathcal{Y} \setminus \{\mathbf{x}\}\}$. Then $\mathcal{O}_\mathbf{x}(\mu) \neq \emptyset$ (because $\mu \in \Delta^*(\mathcal{Y})$), and $\mathcal{O}_\mathbf{x}(\mu)$ is an open subset of $\Delta(\mathcal{Y})$, and μ is a cluster point of $\mathcal{O}_\mathbf{x}(\mu)$. Furthermore, $F(\mu') = \{\mathbf{x}\}$ for all $\mu' \in \mathcal{O}_\mathbf{x}(\mu)$ (because F is monotone). Thus, F is single-valued on $\mathcal{O}_\mathbf{x}(\mu)$.

Next, for any $\mu \in \Delta^*(\mathcal{Y})$, define $\mathcal{O}(\mu) := \displaystyle\bigcup_{\mathbf{y}\in F(\mu)} \mathcal{O}_\mathbf{y}(\mu)$. Then $\mathcal{O}(\mu)$ is a nonempty open subset of $\Delta(\mathcal{Y})$, and μ is a cluster point of $\mathcal{O}(\mu)$ (because it is a union of one or more nonempty open subsets of $\Delta(\mathcal{Y})$ clustering at μ). Finally, let $\mathcal{O} := \displaystyle\bigcup_{\mu\in\Delta^*(\mathcal{Y})} \mathcal{O}(\mu)$. Then $\mathcal{O}$ is an open subset of $\Delta(\mathcal{Y})$, and F is single-valued on $\mathcal{O}$. Every element of $\Delta^*(\mathcal{Y})$ is a cluster point of $\mathcal{O}$, and $\Delta^*(\mathcal{Y})$ is dense in $\Delta(\mathcal{Y})$; thus, $\mathcal{O}$ is dense in $\Delta(\mathcal{Y})$. □

Proof of Proposition 4.4 (a) Suppose $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. Let $\mu \in \Delta(\mathcal{Y})$ and $\mathbf{x} \in F_\phi(\mu)$. Let $\mu' \in \Delta(\mathcal{Y})$ be more supportive than μ of $\mathbf{x}$; we must show that $F_\phi(\mu') = \{\mathbf{x}\}$. By negating certain coordinates of $\mathcal{X}$ if necessary, we can assume without loss of generality that $\mathbf{x} = \mathbf{1}$. Recall that $\mathcal{Y}(\mu) := \{\mathbf{y} \in \mathcal{Y};\ \ \mu(\mathbf{y}) > 0\}$. Thus, $\mathcal{Y}(\mu') \subseteq \mathcal{Y}(\mu)$ (because μ' is more supportive of $\mathbf{x}$). Define $\mathcal{K}_1 := \{k \in \mathcal{K};\ \ \widetilde{\mu}_k = 1\} = \{k \in \mathcal{K};\ \ y_k = 1$ for all $\mathbf{y} \in \mathcal{Y}(\mu)\}$. Let $\mathcal{K}_0 := \mathcal{K} \setminus \mathcal{K}_1$.

Claim 1: (a) *For all $k \in \mathcal{K}$, we have $\widetilde{\mu}'_k \geq \widetilde{\mu}_k$.* (b) *If $k \in \mathcal{K}_0$ then $\widetilde{\mu}'_k > \widetilde{\mu}_k$.*

Proof: (a) If μ' is more supportive than μ of $\mathbf{x}$, then $\mathcal{Y}(\mu') \subseteq \mathcal{Y}(\mu)$. Let $\mathcal{Y}_- := \{\mathbf{y} \in \mathcal{Y}(\mu);\ y_k = -1\}$ and $\mathcal{Y}_+ := \{\mathbf{y} \in \mathcal{Y}(\mu);\ \mathbf{y} \neq \mathbf{x} \text{ but } y_k = 1\}$. Then for all $k \in \mathcal{K}$, we have

$$\begin{aligned}
(\widetilde{\mu}'_k - \widetilde{\mu}_k) &\underset{(\diamond)}{=} \sum_{\mathbf{y} \in \mathcal{Y}(\mu)} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big)\, y_k \\
&= \Big(\mu'(\mathbf{x}) - \mu(\mathbf{x})\Big)\, x_k + \sum_{\mathbf{y} \in \mathcal{Y}_-} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big)\, y_k + \sum_{\mathbf{y} \in \mathcal{Y}_+} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big)\, y_k. \\
&\underset{(\dagger)}{=} \Big(\mu'(\mathbf{x}) - \mu(\mathbf{x})\Big) - \sum_{\mathbf{y} \in \mathcal{Y}_-} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big) + \sum_{\mathbf{y} \in \mathcal{Y}_+} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big) \\
&\underset{(*)}{\geq} \Big(\mu'(\mathbf{x}) - \mu(\mathbf{x})\Big) + \sum_{\mathbf{y} \in \mathcal{Y}_-} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big) + \sum_{\mathbf{y} \in \mathcal{Y}_+} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big) \\
&= \sum_{\mathbf{y} \in \mathcal{Y}(\mu)} \Big(\mu'(\mathbf{y}) - \mu(\mathbf{y})\Big) \quad = \quad 1 - 1 \quad = \quad 0,
\end{aligned}$$

and thus, $\widetilde{\mu}'_k \geq \widetilde{\mu}_k$. Here, $(\diamond)$ is by defining equation (4.1), and the fact that $\mathcal{Y}(\mu') \subseteq \mathcal{Y}(\mu)$. Meanwhile, $(\dagger)$ is by definition of $\mathcal{Y}_-$ and $\mathcal{Y}_+$, and $(*)$ is because $\mu'(\mathbf{y}) - \mu(\mathbf{y}) < 0$ for all $\mathbf{y} \in \mathcal{Y}_-$ by the hypothesis on μ'.

(b) If $k \in \mathcal{K}_0$, then $\mathcal{Y}_- \neq \emptyset$, so $(*)$ becomes a strict inequality, so $\widetilde{\mu}'_k > \widetilde{\mu}_k$. $\diamond$ `Claim 1`

Now, let $\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{x}\}$; we will show that $\mathbf{y} \notin F(\mu')$. There are two cases.

Case 1. Suppose $\mathcal{K}(\mathbf{x}, \mathbf{y}) \subseteq \mathcal{K}_1$. Then for all $k \in \mathcal{K}(\mathbf{x}, \mathbf{y})$, we have $\widetilde{\mu}_k = x_k = 1$, while $y_k = -1$. Thus,

$$(\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\mu}_k) \quad = \quad \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, (x_k - y_k)\, \phi(\widetilde{\mu}_k) \quad = \quad 2\, d(\mathbf{x}, \mathbf{y}) \cdot \phi(1) \quad > \quad 0.$$

Thus, $\mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) > \mathbf{y} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}})$, so $\mathbf{y} \notin F_\phi(\mu)$.

Case 2. Suppose $\mathcal{K}(\mathbf{x}, \mathbf{y}) \not\subseteq \mathcal{K}_1$. For all $k \in \mathcal{K}(\mathbf{x}, \mathbf{y})$, we have $x_k = 1$ and $y_k = -1$, while Claim 1(a) says $\widetilde{\mu}'_k \geq \widetilde{\mu}_k$. Furthermore, $\mathcal{K}_0 \cap \mathcal{K}(\mathbf{x}, \mathbf{y}) \neq \emptyset$, and for any $k \in \mathcal{K}_0 \cap \mathcal{K}(\mathbf{x}, \mathbf{y})$, Claim 1(b) says $\widetilde{\mu}'_k > \widetilde{\mu}_k$. Thus,

$$(\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}') \;=\; \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, 2\phi(\widetilde{\mu}'_k) \underset{(\diamond)}{>} \sum_{k \in \mathcal{K}(\mathbf{x},\mathbf{y})} \lambda_k\, 2\phi(\widetilde{\mu}_k) \;=\; (\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) \underset{(*)}{\geq} 0,$$

and thus, $\mathbf{y} \notin F_\phi(\mu')$. Here, $(\diamond)$ is because ϕ is strictly increasing, and $(*)$ is because $\mathbf{x} \in F_\phi(\mu)$. We conclude that $\mathbf{y} \notin F_\phi(\mu')$ for all $\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{x}\}$; thus, $F(\widetilde{\boldsymbol{\mu}}') = \{\mathbf{x}\}$, as desired.

(b) follows immediately from part (a) and Lemma B.1, because any additive majority rule is total. $\square$

C Preparatory results for Section 5

Theorem 3 is a special case of a more general result, which is in fact a preliminary step in the proof of Theorem 1(b,c,d). To state it, we need some notation.

As explained in Section 3, a profile μ in $\Delta(\mathcal{Y})$ represents a population of weighted voters. To be precise, let $\mathcal{N}$ be a finite set of voters, and let $\omega : \mathcal{N} \longrightarrow \mathbb{R}_+$ be a "weight function" (reflecting, e.g. the differing expertise or standing of different voters). For all $n \in \mathcal{N}$, let $\mathbf{x}^n \in \mathcal{Y}$ be the view of voter n. Define the function $\mu : \mathcal{Y} \longrightarrow [0,1]$ by setting

$$\mu(\mathbf{y}) \quad = \quad \frac{1}{W} \sum \{\omega(n) \; ; \; n \in \mathcal{N} \text{ and } \mathbf{x}^n = \mathbf{y}\}, \qquad \text{for all } \mathbf{y} \in \mathcal{Y}, \tag{C1}$$

where $W := \sum_{n \in \mathcal{N}} \omega(n)$. This function summarizes the voters' views, but renders them anonymous (except for their differing weights).

Now, let $\wp[0,1]$ be the set of all finite subsets of $[0,1]$. For any $\mathcal{H} \in \wp[0,1]$, we define $\Delta_{\mathcal{H}}(\mathcal{Y}) := \{\mu \in \Delta(\mathcal{Y}); \;\; |\widetilde{\mu}_k| \in \mathcal{H} \text{ for all } k \in \mathcal{K}\}$. For example, suppose $\mathcal{H} = \{\frac{m}{N};\; m \in [0 \ldots N]\}$; then $\Delta_{\mathcal{H}}(\mathcal{Y})$ is contains all profiles which can be constructed using exactly N equally weighted voters. More generally, let $\mathcal{N}$ be a finite set of voters, let $\omega : \mathcal{N} \longrightarrow \mathbb{R}_+$ be a weight function, and let $\Delta_\omega(\mathcal{Y})$ be the set of all profiles in $\Delta(\mathcal{Y})$ obtained by using ω in equation (C1). Let $\mathcal{H}$ to be the set of all elements of $[0,1]$ which can be written in the form $\sum_{n \in \mathcal{N}} x_n \, \omega(n)$ for some $\mathbf{x} \in \{\pm 1\}^{\mathcal{K}}$; then $\Delta_\omega(\mathcal{Y}) \subseteq \Delta_{\mathcal{H}}(\mathcal{Y})$.

If $\mathfrak{E}$ is a judgement environment and $\mathcal{H} \in \wp[0,1]$, then we define $\Delta_{\mathcal{H}}(\mathfrak{E})$ to be the set of all judgement probems $(\boldsymbol{\mathcal{C}}, \mu)$ where $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$. For any judgement environment $\mathfrak{G}$, we define $\langle \mathfrak{G} \rangle$ as in formula (5.1).

Theorem C.1 *Let $\mathfrak{G}$ be a finite, rationally weighted judgement environment, let $\mathcal{H} \in \wp[0,1]$, and let F be a judgement aggregation rule on $\Delta_{\mathcal{H}}\langle \mathfrak{G} \rangle$ which satisfies* Combination*.*

- **(a)** *F satisfies* Supermajority efficiency *on $\Delta_{\mathcal{H}}\langle \mathfrak{G} \rangle$ if and only if there is a gain function $\phi_{\mathcal{H}} : \mathcal{H} \longrightarrow \mathbb{R}$ such that $F(\boldsymbol{\mathcal{C}}; \mu) \subseteq F_{\phi_{\mathcal{H}}}(\boldsymbol{\mathcal{C}}; \mu)$ for all $\boldsymbol{\mathcal{C}} \in \langle \mathfrak{G} \rangle$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$.*

- **(b)** *If F is also total, then F has a unique minimal covering on $\Delta_{\mathcal{H}}\langle \mathfrak{G} \rangle$ amongst all additive majority rules.*[28]

- **(c)** *Any additive majority rule G on $\Delta_{\mathcal{H}}\langle \mathfrak{G} \rangle$ has a real-valued representation.*

Theorem C.1 only applies to rationally weighted judgement environments. We will also need a variant of Theorem C.1 for *real*-weighted judgement environments. A (real-weighted) judgement environment $\mathfrak{E}$ is ***finitely generated*** if there is a finite subset $\mathfrak{G} \subseteq \mathfrak{E}$ such that:

(FG1) If $\boldsymbol{\mathcal{C}}_1, \boldsymbol{\mathcal{C}}_2 \in \mathfrak{E}$, then $\boldsymbol{\mathcal{C}}_1 \times \boldsymbol{\mathcal{C}}_2 \in \mathfrak{E}$ also.

(FG2) For all $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$, we have $\boldsymbol{\mathcal{C}} = r_1 \boldsymbol{\mathcal{G}}_1 \times \cdots \times r_J \boldsymbol{\mathcal{G}}_J$ for some $J \in \mathbb{N}$, $\boldsymbol{\mathcal{G}}_1, ..., \boldsymbol{\mathcal{G}}_J \in \mathfrak{G}$, and $r_1, ..., r_J \in \mathbb{R}_+$.

(FG3) For all $\boldsymbol{\mathcal{G}} \in \mathfrak{G}$, there is some $\epsilon > 0$ such that $r\,\boldsymbol{\mathcal{G}} \in \mathfrak{E}$ for all $r \in \mathbb{R}$ with $1 - \epsilon < r < 1 + \epsilon$.

[28]However, the gain function which defines the minimal covering rule is not unique.

In the notation introduced in formula (5.1), condition (FG1) says that $\mathfrak{E} = \langle\mathfrak{E}\rangle$. If we define $\mathfrak{G}' = \{r\boldsymbol{\mathcal{G}};\ \boldsymbol{\mathcal{G}} \in \mathfrak{G} \text{ and } r \in \mathbb{R}_+\}$, then condition (FG2) says that $\mathfrak{E} \subseteq \langle\mathfrak{G}'\rangle$. Finally, condition (FG3) is just the property of *minimal richness* defined prior to Axiom 3. Note that we can use the same ϵ for all $\boldsymbol{\mathcal{G}}$ in $\mathfrak{G}$, because $\mathfrak{G}$ is finite. For example, let $\mathfrak{G}$ be a finite collection of judgement contexts, let $0 < \underline{r} < 1 < \overline{r} \leq \infty$, and define $\langle\mathfrak{G}, \underline{r}, \overline{r}\rangle := \{r_1 \boldsymbol{\mathcal{G}}_1 \times \cdots \times r_N \boldsymbol{\mathcal{G}}_N;\ N \in \mathbb{N},\ \ \underline{r} < r_1, \ldots, r_N < \overline{r} \text{ and } \boldsymbol{\mathcal{G}}_1, \ldots, \boldsymbol{\mathcal{G}}_N \in \mathfrak{G}\}$; then $\langle\mathfrak{G}, \underline{r}, \overline{r}\rangle$ is a finitely generated judgement environment.

Theorem C.2 *Let $\mathcal{H} \in \wp[0,1]$, let $\mathfrak{E}$ be a finitely generated judgement environment, and let F be a judgement aggregation rule on $\Delta_{\mathcal{H}}(\mathfrak{E})$ which satisfies* Combination *and* Scale invariance. *Then the statements of Theorem C.1*(a,b,c) *are true if we replace every instance of "$\langle\mathfrak{G}\rangle$" with "$\mathfrak{E}$".*

The advantage of Theorem C.2 over Theorem C.1 is that Theorem C.2 applies *any* judgement context (as long as it is finitely generated), whereas Theorem C.1 only applies to finite, rationally-weighted contexts. By contrast, the advantage of Theorem C.1 over Theorem C.2 is that Theorem C.1 does not require $\mathfrak{G}$ to satisfy the richness condition (FG3). It also does not require F to satisfy Scale invariance. Surprisingly, for *rationally*-weighted judgement contexts, the axioms of Supermajority Efficiency and Combination alone are sufficient to yield a covering by an additive majority rule.

For the rest of this appendix, let $\mathcal{H}$, $\mathfrak{E}$ and F be as in Theorem C.2. Sometimes, we will use the symbol $\mathbb{P}$ to represent "either $\mathbb{Q}$ or $\mathbb{R}$".[29] For any $J \in \mathbb{N}$, let $\Delta^J_{\mathbb{P}} := \{(p_1, \ldots, p_J) \in \mathbb{P}^J_+;\ \ p_1 + \cdots + p_J = 1\}$. For any subset $\mathcal{V} \subseteq \mathbb{R}^{\mathcal{H}}$, let $\mathrm{conv}_{\mathbb{P}}(\mathcal{V})$ denote the $\mathbb{P}$-*convex hull* of $\mathcal{V}$. Formally, $\mathrm{conv}_{\mathbb{P}}(\mathcal{V}) := \{\sum_{j=1}^{J} p_j\, \mathbf{v}_j;\ \ J \in \mathbb{N},\ \mathbf{v}_1, \ldots, \mathbf{v}_J \in \mathcal{V}, \text{ and } (p_1, \ldots, p_J) \in \Delta^J_{\mathbb{P}}\}$. Clearly, $\mathrm{conv}_{\mathbb{P}}(\mathcal{V}) \subset \mathbb{R}^{\mathcal{H}}$. Furthermore, if $\mathcal{V} \subset \mathbb{P}^{\mathcal{H}}$, then $\mathrm{conv}_{\mathbb{P}}(\mathcal{V}) \subset \mathbb{P}^{\mathcal{H}}$. Let $\mathbb{R}^{\mathcal{H}}_+$ be the non-negative orthant in $\mathbb{R}^{\mathcal{H}}$, while $\mathbb{R}^{\mathcal{H}}_{++}$ is the strictly positive orthant in $\mathbb{R}^{\mathcal{H}}$. Formally, $\mathbb{R}^{\mathcal{H}}_+ := \{\mathbf{r} \in \mathbb{R}^{\mathcal{H}};\ \ r_h \geq 0 \text{ for all } h \in \mathcal{H}\}$ and $\mathbb{R}^{\mathcal{H}}_{++} := \{\mathbf{r} \in \mathbb{R}^{\mathcal{H}};\ \ r_h > 0 \text{ for all } h \in \mathcal{H}\}$. The proof of Theorems C.1 and C.2 occupies the rest of the appendix, and involves seven preliminary lemmas. The first one strengthens a well-known result about Pareto optima in convex sets.

Lemma C.1 *Let $\mathbb{P} = \mathbb{Q}$ or $\mathbb{R}$. Let $\mathcal{V} \subseteq \mathbb{P}^{\mathcal{H}}$ be a finite set, let $\mathcal{P} := \mathrm{conv}_{\mathbb{P}}(\mathcal{V})$, and let $\overline{\mathcal{P}} := \mathrm{conv}(\mathcal{V})$. If $\mathcal{P} \cap \mathbb{R}^{\mathcal{H}}_+ = \{\mathbf{0}\}$, then there exists $\mathbf{z} \in \mathbb{R}^{\mathcal{H}}_{++}$ such that $\mathbf{z} \bullet \mathbf{p} \leq 0$ for all $\mathbf{p} \in \overline{\mathcal{P}}$.*

Proof:

Claim 1: *If $\mathcal{P} \cap \mathbb{R}^{\mathcal{H}}_+ = \{\mathbf{0}\}$, then $\overline{\mathcal{P}} \cap \mathbb{R}^{\mathcal{H}}_+ = \{\mathbf{0}\}$.*

Proof: (by contrapositive) Clearly, $\mathbf{0} \in \overline{\mathcal{P}} \cap \mathbb{R}^{\mathcal{H}}_+$ because $\mathbf{0} \in \mathcal{P} \cap \mathbb{R}^{\mathcal{H}}_+$. Suppose $\overline{\mathcal{P}} \cap \mathbb{R}^{\mathcal{H}}_+ \neq \{\mathbf{0}\}$; we will show that $\mathcal{P} \cap \mathbb{R}^{\mathcal{H}}_+ \neq \{\mathbf{0}\}$ also.

[29] In fact, for all the proofs in this appendix —including the proof of Theorem C.2 —we could replace $\mathbb{R}$ with an arbitrary subfield of $\mathbb{R}$. In this case, $\mathbb{P}$ can be taken to represent this subfield. However, this additional level of generality is not necessary for the proofs of Theorem 1, 2, or 3, so we have not made it explicit in this appendix.

Let $\mathbf{p} \in \overline{\mathcal{P}} \cap \mathbb{R}^{\mathcal{H}}_{+}$ be nonzero. Then $p_h \geq 0$ for all $h \in \mathcal{H}$, and $p_k > 0$ for some $k \in \mathcal{H}$. Suppose $\mathcal{V} = \{\mathbf{v}^1, \ldots, \mathbf{v}^J\}$, where $\mathbf{v}^j = (v_1^j, \ldots, v_H^j)$ for all $j \in [1 \ldots J]$. Consider the following linear programming problem:

$$\text{maximize} \quad \sum_{j=1}^{J} r_j \, v_k^j \qquad \text{over all } (r_1, \ldots, r_J) \in \mathbb{R}^J, \tag{C2}$$

$$\text{subject to} \quad r_j \; \geq \; 0, \quad \text{for all } j \in [1 \ldots J]; \tag{C3}$$

$$\sum_{j=1}^{J} r_j \; = \; 1; \tag{C4}$$

$$\text{and} \quad \sum_{j=1}^{J} r_j \, v_h^j \; \geq \; 0, \quad \text{for all } h \in \mathcal{H}. \tag{C5}$$

Since $\mathbf{p} \in \overline{\mathcal{P}}$, we know that $\mathbf{p} = \sum_{j=1}^{J} r_j \, \mathbf{v}^j$ for some $\mathbf{r} = (r_1, \ldots, r_J) \in \Delta^J$ — equivalently, some $\mathbf{r} \in \mathbb{R}^J$ satisfying the constraints (C3) and (C4). But $\mathbf{p} \in \mathbb{R}^{\mathcal{H}}_{+}$, so $\mathbf{r}$ also satisfies constraint (C5). Finally, $p_k > 0$, so the optimal value of objective function (C2) is strictly positive.

Now, suppose we solve the linear program (C2)-(C5) using the Simplex Method. Note that the objective function (C2) and the constraints (C3)-(C5) are all $\mathbb{P}$-linear. Thus, the output of the Simplex Method will also be $\mathbb{P}$-linear (because each pivot operation converts $\mathbb{P}$-valued input into $\mathbb{P}$-valued output). Thus, we obtain an optimal solution $\mathbf{s} \in \mathbb{P}^{\mathcal{H}}$. Let $\mathbf{q} = \sum_{j=1}^{J} s_j \, \mathbf{v}^j$. Constraints (C3)-(C4) imply that $\mathbf{s} \in \Delta^J_{\mathbb{P}}$; thus $\mathbf{q} \in \mathcal{P}$. Meanwhile, constraint (C5) says that $\mathbf{q} \in \mathbb{R}^{\mathcal{H}}_{+}$. Finally, $q_k \geq p_k > 0$; thus, $\mathbf{q} \neq \mathbf{0}$, as claimed. $\diamond$ `Claim 1`

Note that $\overline{\mathcal{P}}$ is a compact, convex polyhedron in $\mathbb{R}^{\mathcal{H}}$. Let $\mathcal{T} := \overline{\mathcal{P}} - \mathbb{R}^{\mathcal{H}}_{+}$; then $\mathcal{T}$ is also a closed, convex polyhedron in $\mathbb{R}^{\mathcal{H}}$.

Claim 2: $\mathcal{T} \cap \mathbb{R}^{\mathcal{H}}_{+} = \{\mathbf{0}\}$.

Proof: (by contradiction) If $\mathbf{t} \in \mathcal{T}$, then $\mathbf{t} = \mathbf{p} - \mathbf{r}$ for some $\mathbf{p} \in \overline{\mathcal{P}}$ and $\mathbf{r} \in \mathbb{R}^{\mathcal{H}}_{+}$. Thus $\mathbf{p} = \mathbf{t} + \mathbf{r}$. Thus, if $\mathbf{t} \in \mathbb{R}^{\mathcal{H}}_{+} \setminus \{\mathbf{0}\}$, then $\mathbf{p} \in \mathbb{R}^{\mathcal{H}}_{+} \setminus \{\mathbf{0}\}$ also, contradicting Claim 1. $\diamond$ `Claim 2`

Thus, $\mathbf{0}$ lies on the boundary of the polyhedron $\mathcal{T}$, so it is contained in some face. Let $\mathcal{F}$ be the minimal-dimension face of $\mathcal{T}$ containing $\mathbf{0}$. Thus, if $\mathcal{S} \subseteq \mathbb{R}^{\mathcal{H}}$ is the linear subspace spanned by $\mathcal{F}$, then $\mathcal{F} = \mathcal{S} \cap \mathcal{T}$, and there exists some $\epsilon > 0$ such that $\mathcal{B}_\epsilon \cap \mathcal{S} \subset \mathcal{F}$, where $\mathcal{B}_\epsilon \subset \mathbb{R}^{\mathcal{H}}$ is the ϵ-ball around $\mathbf{0}$.

$\mathcal{F}$ is a face of $\mathcal{T}$, so there exists some $\mathbf{z} \in \mathbb{R}^{\mathcal{K}}$ such that $\mathcal{F} = \operatorname{argmax}_{\mathbf{t} \in \mathcal{T}} (\mathbf{z} \bullet \mathbf{t})$. Clearly $\mathbf{z} \bullet \mathbf{0} = 0$; hence $\mathbf{z} \bullet \mathbf{f} = 0$ for all $\mathbf{f} \in \mathcal{F}$, while $\mathbf{z} \bullet \mathbf{t} < 0$ for all $\mathbf{t} \in \mathcal{T} \setminus \mathcal{F}$. Thus, if $\mathcal{W} \subset \mathbb{R}^{\mathcal{H}}$ is the hyperplane orthogonal to $\mathbf{z}$, then $\mathcal{F} = \mathcal{T} \cap \mathcal{W}$.

Claim 3: $\mathbf{z} \bullet \mathbf{r} > 0$ *for all* $\mathbf{r} \in \mathbb{R}^{\mathcal{H}}_{+} \setminus \{\mathbf{0}\}$.

Proof: For any $\mathbf{r} \in \mathbb{R}_+^{\mathcal{H}} \setminus \{\mathbf{0}\}$, we have $-\mathbf{r} \in \mathcal{T}$, so $\mathbf{z} \bullet (-\mathbf{r}) \leq 0$, so $\mathbf{z} \bullet \mathbf{r} \geq 0$. It remains to show that $\mathbf{z} \bullet \mathbf{r} \neq 0$. By contradiction, suppose $\mathbf{z} \bullet \mathbf{r} = 0$. Then $\mathbf{r} \in \mathcal{W}$. Thus, $-\mathbf{r} \in \mathcal{W}$. But $-\mathbf{r} \in \mathcal{T}$ also, so $-\mathbf{r} \in \mathcal{T} \cap \mathcal{W} = \mathcal{F} = \mathcal{T} \cap \mathcal{S}$. But $\mathcal{F}$ contains a relative neighbourhood around $\mathbf{0}$ in the subspace $\mathcal{S}$, so if $-\mathbf{r} \in \mathcal{F}$, then there exists some $\epsilon > 0$ such that $\epsilon\,\mathbf{r} \in \mathcal{F}$; hence $\epsilon\,\mathbf{r} \in \mathcal{T}$. But $\epsilon\,\mathbf{r} \in \mathbb{R}_+^{\mathcal{H}} \setminus \{\mathbf{0}\}$, so this contradicts Claim 2. ◇ `Claim 3`

For any $h \in \mathcal{H}$, let $\mathbf{e}_h = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ be the hth unit vector. Then $\mathbf{e}_h \in \mathbb{R}_+^{\mathcal{H}}$. Thus, Claim 3 says that $\mathbf{z} \bullet \mathbf{e}_h > 0$. Thus, $z_h > 0$, as desired. □

Let $\mathbb{P} = \mathbb{Q}$ or $\mathbb{R}$, and let $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ be a $\mathbb{P}$-weighted judgement context. For notational convenience, we define $|\mathcal{C}| := |\mathcal{K}|$. Let $\mathbf{x} \in \mathcal{X}$ and let $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$, and define $\gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}} : [0,1] \longrightarrow \mathbb{R}_+$ as in formula (4.2). Note that $\gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}}(q) \in \mathbb{P}_+$ for all $q \in [0,1]$ (because $\mathbb{P}$ is closed under addition). Furthermore, $\gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}}$ is constant on any subinterval of $[0,1]$ not containing an element of $\mathcal{H}$ (by definition of $\Delta_{\mathcal{H}}(\mathcal{Y})$). Thus, a complete description of $\gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}}$ is encoded in the $|\mathcal{H}|$-dimensional vector $\mathbf{g}(\mathbf{x}, \mathcal{C}, \mu) \in \mathbb{P}^{\mathcal{H}}$ defined by setting

$$g_h(\mathbf{x}, \mathcal{C}, \mu) \quad := \quad \frac{1}{|\mathcal{C}|}\gamma_{\mu,\mathbf{x}}^{\boldsymbol{\lambda}}(h) \quad = \quad \frac{1}{|\mathcal{K}|} \sum_{\substack{k \in \mathcal{K} \text{ with} \\ x_k\, \widetilde{\mu}_k \geq h}} \lambda_k, \qquad \text{for all } h \in \mathcal{H}. \tag{C6}$$

Let $\mathcal{D}_{\mathcal{H}}(\mathbf{x}; \mathcal{C}, \mu) := \{\mathbf{g}(\mathbf{y}, \mathcal{C}, \mu) - \mathbf{g}(\mathbf{x}, \mathcal{C}, \mu);\ \ \mathbf{y} \in \mathcal{X}\}$, a finite subset of $\mathbb{P}^{\mathcal{H}}$. (Note that $\mathbf{0} \in \mathcal{D}_{\mathcal{H}}(\mathbf{x}; \mathcal{C}, \mu)$.) If $F : \Delta_{\mathcal{H}}(\mathcal{Y}) \rightrightarrows \mathcal{X}$ is a judgement aggregation rule, then we define

$$\mathcal{D}_{\mathcal{H}}(F, \mathcal{C}) \quad := \quad \bigcup_{\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})} \ \bigcup_{\mathbf{x} \in F(\mathcal{C}, \mu)} \mathcal{D}_{\mathcal{H}}(\mathbf{x}; \mathcal{C}, \mu) \quad \subseteq \quad \mathbb{P}^{\mathcal{H}}. \tag{C7}$$

Finally, let $\mathfrak{E}$ be a $\mathbb{P}$-weighted judgement environment, and let F be a judgement aggregation rule on $\Delta_{\mathcal{H}}(\mathfrak{E})$. Then we define

$$\mathcal{D}_{\mathcal{H}}(F, \mathfrak{E}) \quad := \quad \bigcup_{\mathcal{C} \in \mathfrak{E}} \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}) \quad \subseteq \quad \mathbb{P}^{\mathcal{H}}. \tag{C8}$$

Note that $\mathbf{0} \in \mathcal{D}_{\mathcal{H}}(F, \mathfrak{E})$, because $\mathbf{0} \in \mathcal{D}_{\mathcal{H}}(\mathbf{x}; \mu)$ for any choice of $\mathbf{x}$ and μ. The next lemma shows how the supermajority efficiency of the rule F implies that $\mathcal{D}_{\mathcal{H}}(F, \mathfrak{E})$ satisfies the key hypothesis of Lemma C.1.

Lemma C.2 *F satisfies* Supermajority efficiency *on $\Delta_{\mathcal{H}}(\mathfrak{E})$ if and only if $\mathcal{D}_{\mathcal{H}}(F, \mathfrak{E}) \cap \mathbb{R}_+^{\mathcal{H}} = \{\mathbf{0}\}$.*

Proof: For any $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ in $\mathfrak{E}$, any $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$, and any $\mathbf{x} \in \mathcal{X}$, we have

$$\begin{aligned}
&\Big(\mathbf{x} \in \mathrm{SME}\,(\mathcal{C}, \mu)\Big) \\
&\iff \Big(\nexists\ \mathbf{y} \in \mathcal{X} \text{ with } g_h(\mathbf{y}, \mathcal{C}, \mu) \geq g_h(\mathbf{x}, \mathcal{C}, \mu) \text{ for all } h \in \mathcal{H}, \text{ and } \mathbf{g}(\mathbf{y}, \mathcal{C}, \mu) \neq \mathbf{g}(\mathbf{x}, \mathcal{C}, \mu)\Big) \\
&\iff \Big((\mathbf{g}(\mathbf{y}, \mathcal{C}, \mu) - \mathbf{g}(\mathbf{x}, \mathcal{C}, \mu)) \notin \mathbb{R}_+^{\mathcal{H}} \setminus \{\mathbf{0}\} \text{ for all } \mathbf{y} \in \mathcal{X}\Big) \\
&\iff \Big(\mathcal{D}_{\mathcal{H}}(\mathbf{x}, \mathcal{C}, \mu) \cap \mathbb{R}_+^{\mathcal{H}} = \{\mathbf{0}\}\Big).
\end{aligned}$$

Thus, F satisfies Supermajority efficiency on $\Delta_{\mathcal{H}}(\mathfrak{E})$ if and only if $\mathcal{D}_{\mathcal{H}}(\mathbf{x},\boldsymbol{\mathcal{C}},\mu)\cap\mathbb{R}^{\mathcal{H}}_{+}=\{\mathbf{0}\}$ whenever $\mathbf{x}\in F(\boldsymbol{\mathcal{C}},\mu)$ for some $(\boldsymbol{\mathcal{C}};\mu)\in\Delta_{\mathcal{H}}(\mathfrak{E})$. The claim now follows from defining formulae (C7) and (C8). □

Suppose $\mathfrak{E}$ is a finitely generated judgement context, generated by some finite subset $\mathfrak{G}\subset\mathfrak{E}$. Condition (FG3) implies that there is some $\epsilon=\epsilon(\mathfrak{E})>0$ such that

$$r\,\boldsymbol{\mathcal{C}}\in\mathfrak{E}\text{ for all }\boldsymbol{\mathcal{C}}\in\mathfrak{G}\text{ and all }r\in(1-\epsilon,1+\epsilon). \tag{C9}$$

For any $N\in\mathbb{N}$, and any $K_1,\ldots,K_N\in\mathbb{N}$, we define

$$\mathcal{P}_\epsilon(K_1,...,K_N):=\left\{(p_1,...,p_N)\ ;\ p_1,...,p_N\in(1-\epsilon,1+\epsilon)\text{ and }\ \sum_{n=1}^{N}p_nK_n=\sum_{n=1}^{N}K_n\right\}. \tag{C10}$$

Thus, if $K:=\sum_{n=1}^{N}K_n$ and $q_n:=K_n/K$ for all $n\in[1\ldots N]$, then for any $(p_1,\ldots,p_N)\in\mathcal{P}_\epsilon(K_1,\ldots,K_N)$, we have $q_1\,p_1+\cdots+q_N\,p_N=1$, so that $(q_1\,p_1,\ldots,q_N\,p_N)\in\Delta^N_{\mathbb{P}}$. Let $\mathbf{1}:=(1,1,\ldots,1)$; then $\mathbf{1}\in\mathcal{P}_\epsilon(K_1,\ldots,K_N)$, and $\mathcal{P}_\epsilon(K_1,\ldots,K_N)$ can be visualized as a small neighbourhood of $\mathbf{1}$ in the affine hyperplane of $\mathbb{P}^N$ orthogonal to the vector $(q_1,\ldots,q_N)$. Now let

$$\langle\mathfrak{G},\epsilon\rangle:=\{p_1\boldsymbol{\mathcal{C}}_1\times\cdots\times p_N\boldsymbol{\mathcal{C}}_N\ ;\ N\in\mathbb{N},\ \boldsymbol{\mathcal{C}}_1,...,\boldsymbol{\mathcal{C}}_N\in\mathfrak{G}\text{ and }\ (p_1,...,p_N)\in\mathcal{P}_\epsilon(|\boldsymbol{\mathcal{C}}_1|,...,|\boldsymbol{\mathcal{C}}_N|)\}. \tag{C11}$$

Then conditions (FG1) and (FG3) together imply that $\langle\mathfrak{G},\epsilon\rangle\subseteq\mathfrak{E}$. Thus, if F is a judgement aggregation rule on $\Delta_{\mathcal{H}}(\mathfrak{E})$, then F is well-defined on $\Delta_{\mathcal{H}}\langle\mathfrak{G},\epsilon\rangle$. Note that $\langle\mathfrak{G}\rangle\subseteq\langle\mathfrak{G},\epsilon\rangle$, where $\langle\mathfrak{G}\rangle$ is defined as in formula (5.1). The next lemma shows how the $\mathbb{Q}$-convexity of $\mathcal{D}_{\mathcal{H}}(F,\langle\mathfrak{G}\rangle)$ arises from the axiom of Combination, while the $\mathbb{R}$-convexity of $\mathcal{D}_{\mathcal{H}}(F,\langle\mathfrak{G},\epsilon\rangle)$ arises from Combination and Scale invariance.

Lemma C.3 *Suppose $\mathfrak{G}=\{\boldsymbol{\mathcal{C}}_1,\ldots,\boldsymbol{\mathcal{C}}_J\}$, and let $\mathcal{V}:=\mathcal{D}_{\mathcal{H}}(F,\boldsymbol{\mathcal{C}}_1)\cup\cdots\cup\mathcal{D}_{\mathcal{H}}(F,\boldsymbol{\mathcal{C}}_J)$. Then:*

(a) *$\mathcal{V}$ is finite.*

(b) *If F satisfies* Combination *on $\Delta_{\mathcal{H}}\langle\mathfrak{G}\rangle$, then $\mathcal{D}_{\mathcal{H}}(F,\langle\mathfrak{G}\rangle)=\mathrm{conv}_{\mathbb{Q}}(\mathcal{V})$.*

(c) *If F satisfies* Combination *and* Scale invariance *on $\Delta_{\mathcal{H}}\langle\mathfrak{G},\epsilon\rangle$, then $\mathcal{D}_{\mathcal{H}}(F,\langle\mathfrak{G},\epsilon\rangle)=\mathrm{conv}(\mathcal{V})$.*

The proof of Lemma C.3 depends on another lemma.

Lemma C.4 **(a)** *Let $\mathfrak{G}$ be a $\mathbb{Q}$-weighted judgement environment. If F satisfies* Combination *on $\Delta_{\mathcal{H}}\langle\mathfrak{G}\rangle$, then for all $J\in\mathbb{N}$ and all $\boldsymbol{\mathcal{C}}^1,\ldots,\boldsymbol{\mathcal{C}}^J\in\langle\mathfrak{G}\rangle$,*

$$\mathcal{D}_{\mathcal{H}}(F,\boldsymbol{\mathcal{C}}^1\times\cdots\times\boldsymbol{\mathcal{C}}^J)\ =\ \left\{q_1\,\mathbf{d}^1+\cdots+q_J\,\mathbf{d}^J\ ;\ \mathbf{d}^j\in\mathcal{D}_{\mathcal{H}}(F,\boldsymbol{\mathcal{C}}^j)\text{ for all }j\in[1\ldots J]\right\}.$$

where $q_j:=|\boldsymbol{\mathcal{C}}_j|/|\boldsymbol{\mathcal{C}}|$ for all $j\in[1\ldots J]$.

(b) *Let $\mathbb{P} = \mathbb{Q}$ or $\mathbb{R}$, and let $\mathfrak{E}$ be a $\mathbb{P}$-weighted judgement environment. If F satisfies* Combination *and* Scale invariance *on $\Delta_{\mathcal{H}}(\mathfrak{E})$, then for all $J \in \mathbb{N}$, all $\mathcal{C}^1, \ldots, \mathcal{C}^J \in \mathfrak{E}$ and all $(p_1, \ldots, p_J) \in \mathbb{P}^J_+$, if $p_1 \mathcal{C}^1 \times \cdots \times p_J \mathcal{C}^J \in \mathfrak{E}$, then*

$$\mathcal{D}_{\mathcal{H}}(F, p_1 \mathcal{C}^1 \times \cdots \times p_J \mathcal{C}^J) = \big\{q_1\, p_1\, \mathbf{d}^1 + \cdots + q_J\, p_J\, \mathbf{d}^J \ ; \ \mathbf{d}^j \in \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}^j) \text{ for all } j \in [1 \ldots J]\big\}.$$

where $q_j := |\mathcal{C}_j|/|\mathcal{C}|$ for all $j \in [1 \ldots J]$.

Proof: Part (a) follows from part (b) in the special case when $\mathbb{P} = \mathbb{Q}$ and $p_1 = \cdots = p_J = 1$. So it suffices to prove the more general part (b), and observe that, when $p_1 = \cdots = p_J = 1$, the following proof does not require F to satisfy Scale invariance; it only requires F to satisfy Combination on $\langle \mathfrak{G} \rangle$.

For any $\mathcal{C} \in \langle \mathfrak{E}, \mathbb{P} \rangle$ and $\mu \in \Delta_{\mathcal{H}}(\mathcal{C})$, define

$$\mathcal{D}_{\mathcal{H}}(F, \mathcal{C}, \mu) \ := \ \bigcup_{\mathbf{x} \in F(\mathcal{C}, \mu)} \mathcal{D}_{\mathcal{H}}(\mathbf{x}; \mathcal{C}, \mu), \tag{C12}$$

$$\text{so that} \quad \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}) \ = \ \bigcup_{\mu \in \Delta_{\mathcal{H}}(\mathcal{C})} \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}, \mu). \tag{C13}$$

Claim 1: *Let $\mathcal{C}^1 = (\mathcal{K}^1, \boldsymbol{\lambda}^1, \mathcal{X}^1, \mathcal{Y}^1)$ and $\mathcal{C}^2 = (\mathcal{K}^2, \boldsymbol{\lambda}^2, \mathcal{X}^2, \mathcal{Y}^2)$ be two judgement contexts in $\mathfrak{E}$, let $(p_1, p_2) \in \mathbb{P}^2_+$, and let $p_1 \mathcal{C}^1 \times p_2 \mathcal{C}^2 = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. (So $\mathcal{K} = \mathcal{K}^1 \sqcup \mathcal{K}^2$, $\mathcal{X} = \mathcal{X}^1 \times \mathcal{X}^2$, $\mathcal{Y} = \mathcal{Y}^1 \times \mathcal{Y}^2$, and $\boldsymbol{\lambda} = (p_1 \boldsymbol{\lambda}^1, p_2 \boldsymbol{\lambda}^2)$.) Suppose $p_1 \mathcal{C}^1 \times p_2 \mathcal{C}^2 \in \mathfrak{E}$. Then*

(a) *For any $\mu_1 \in \Delta_{\mathcal{H}}(\mathcal{Y}^1)$ and $\mu_2 \in \Delta_{\mathcal{H}}(\mathcal{Y}^2)$, there exists a profile $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$ such that $\mu^{(1)} = \mu_1$ and $\mu^{(2)} = \mu_2$, where these marginals are defined as in equation (5.3).*

(b) *Let $q_1 := |\mathcal{K}^1|/|\mathcal{K}|$ and $q_2 := |\mathcal{K}^2|/|\mathcal{K}|$. Then for any $\mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2) \in \mathcal{X}$, we have $\mathbf{g}(\mathbf{x}, \mathcal{C}, \mu) \ = \ q_1\, p_1\, \mathbf{g}(\mathbf{x}^1, \mathcal{C}^1, \mu^{(1)}) + q_2\, p_2\, \mathbf{g}(\mathbf{x}^2, \mathcal{C}^2, \mu^{(2)})$.*

(c) *$\mathcal{D}_{\mathcal{H}}(F, \mathcal{C}, \mu) \ = \ \{q_1\, p_1\, \mathbf{d}^1 + q_2\, p_2\, \mathbf{d}^2; \ \mathbf{d}^1 \in \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}^1, \mu^{(1)})$ and $\mathbf{d}^2 \in \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}^2, \mu^{(2)})\}$.*

(d) *$\mathcal{D}_{\mathcal{H}}(F, \mathcal{C}) = \{q_1\, p_1\, \mathbf{d}^1 + q_2\, p_2\, \mathbf{d}^2; \ \mathbf{d}^1 \in \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}^1)$ and $\mathbf{d}^2 \in \mathcal{D}_{\mathcal{H}}(F, \mathcal{C}^2\}$.*

Proof: (a) Let $\mu := \mu_1 \otimes \mu_2$ —in other words, $\mu(\mathbf{x}^1, \mathbf{x}^2) = \mu_1(\mathbf{x}^1) \cdot \mu_2(\mathbf{x}^2)$, for any $(\mathbf{x}^1, \mathbf{x}^2) \in \mathcal{Y}^1 \times \mathcal{Y}^2$. Then for any $\mathbf{x}^1 \in \mathcal{Y}^1$, we have

$$\mu^{(1)}(\mathbf{x}^1) \ \underset{(*)}{=} \ \sum_{\mathbf{x}^2 \in \mathcal{Y}^2} \mu(\mathbf{x}^1, \mathbf{x}^2) \quad = \quad \mu_1(\mathbf{x}^1) \cdot \sum_{\mathbf{x}^2 \in \mathcal{Y}^2} \mu_2(\mathbf{x}^2) \quad = \quad \mu_1(\mathbf{x}^1),$$

where $(*)$ is by equation (5.3). Thus, $\mu^{(1)} = \mu_1$. Thus, for any $k \in \mathcal{K}^1$, we have $|\widetilde{\mu}_k| = |(\widetilde{\boldsymbol{\mu}}_1)_k| \in \mathcal{H}$. Likewise, $\mu^{(2)} = \mu_2$, and for any $k \in \mathcal{K}^2$, we have $|\widetilde{\mu}_k| = |(\widetilde{\boldsymbol{\mu}}_2)_k| \in \mathcal{H}$. But $\mathcal{K} = \mathcal{K}^1 \sqcup \mathcal{K}^2$. Thus, $|\widetilde{\mu}_k| \in \mathcal{H}$ for all $k \in \mathcal{K}$, so $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$, as desired.

(b) Note that $\widetilde{\boldsymbol{\mu}} = (\widetilde{\boldsymbol{\mu}}^{(1)}, \widetilde{\boldsymbol{\mu}}^{(2)})$. Thus, for both $n \in \{1,2\}$ and all $k \in \mathcal{K}^n$, we have $x_k \cdot \widetilde{\mu}_k = x_k^n \cdot \widetilde{\mu}_k^{(n)}$. Meanwhile, $\mathcal{K} = \mathcal{K}^1 \sqcup \mathcal{K}^2$. Thus, for any $h \in \mathcal{H}$, we have

$$\{k \in \mathcal{K} \,;\; x_k \cdot \widetilde{\mu}_k \geq h\} \quad = \quad \Big\{k \in \mathcal{K}^1 \,;\; x_k^1 \cdot \widetilde{\mu}_k^{(1)} \geq h\Big\} \sqcup \Big\{k \in \mathcal{K}^2 \,;\; x_k^2 \cdot \widetilde{\mu}_k^{(2)} \geq h\Big\}.$$

Thus,
$$\begin{aligned} g_h(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu) &\underset{(*)}{=} \frac{1}{|\mathcal{K}|} \sum \{\lambda_k \,;\; k \in \mathcal{K} \text{ and } x_k \cdot \widetilde{\mu}_k \geq h\} \\ &\underset{(\diamond)}{=} \frac{q_1}{|\mathcal{K}^1|} \sum \Big\{\lambda_k \,;\; k \in \mathcal{K}^1 \text{ and } x_k^1 \cdot \widetilde{\mu}_k^{(1)} \geq h\Big\} + \frac{q_2}{|\mathcal{K}^2|} \sum \Big\{\lambda_k \,;\; k \in \mathcal{K}^2 \text{ and } x_k^2 \cdot \widetilde{\mu}_k^{(2)} \geq h\Big\} \\ &\underset{(\dagger)}{=} \frac{q_1}{|\mathcal{K}^1|} \sum \Big\{p_1\, \lambda_k^1 \,;\; k \in \mathcal{K}^1 \text{ and } x_k^1 \cdot \widetilde{\mu}_k^{(1)} \geq h\Big\} + \frac{q_2}{|\mathcal{K}^2|} \sum \Big\{p_2\, \lambda_k^2 \,;\; k \in \mathcal{K}^2 \text{ and } x_k^2 \cdot \widetilde{\mu}_k^{(2)} \geq h\Big\} \\ &\underset{(*)}{=} q_1\, p_1\, g_h(\mathbf{x}^1, \boldsymbol{\mathcal{C}}^1, \mu^{(1)}) + q_2\, p_2\, g_h(\mathbf{x}^2, \boldsymbol{\mathcal{C}}^1 \mu^{(2)}), \end{aligned}$$

as claimed. Here, $(\diamond)$ is because $q_1 = |\mathcal{K}^1|/|\mathcal{K}|$ and $q_2 = |\mathcal{K}^2|/|\mathcal{K}|$, both $(*)$ are by equation (C6), while $(\dagger)$ is because $\boldsymbol{\lambda} = (p_1 \boldsymbol{\lambda}^1, p_2 \boldsymbol{\lambda}^2)$.

(c) Let $\mathbf{x} := (\mathbf{x}^1, \mathbf{x}^2)$ and $\mathbf{y} := (\mathbf{y}^1, \mathbf{y}^2)$ be elements of $\mathcal{X}$. Thus, $\mathbf{x}^1, \mathbf{y}^1 \in \mathcal{X}^1$ and $\mathbf{x}^2, \mathbf{y}^2 \in \mathcal{X}^2$. If $\mathbf{d} = \mathbf{g}(\mathbf{y}, \mu) - \mathbf{g}(\mathbf{x}, \mu)$, and $\mathbf{d}^n = \mathbf{g}(\mathbf{y}^n, \mu^{(n)}) - \mathbf{g}(\mathbf{x}^n, \mu^{(n)})$ for $n \in \{1,2\}$, then part (b) implies that

$$\begin{aligned} \mathbf{d} &= \mathbf{g}(\mathbf{y}, \mu) - \mathbf{g}(\mathbf{x}, \mu) \\ &= q_1\, p_1\, \mathbf{g}(\mathbf{y}^1, \mu^{(1)}) - q_1\, p_1\, \mathbf{g}(\mathbf{x}^1, \mu^{(1)}) + q_2\, p_2\, \mathbf{g}(\mathbf{y}^2, \mu^{(2)}) - q_2\, p_2\, \mathbf{g}(\mathbf{x}^2, \mu^{(2)}) \\ &= q_1\, p_1\, \mathbf{d}^1 + q_2\, p_2\, \mathbf{d}^2. \end{aligned} \tag{C14}$$

But for any $\mathbf{d} \in \mathbb{R}^{\mathcal{H}}$, we have:

$$\begin{aligned} \Big(\mathbf{d} \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}, \mu)\Big) &\iff \Big(\exists\, \mathbf{x} \in F(\boldsymbol{\mathcal{C}}, \mu) \text{ and } \mathbf{y} \in \mathcal{X} \text{ such that } \mathbf{d} = \mathbf{g}(\mathbf{y}, \mu) - \mathbf{g}(\mathbf{x}, \mu)\Big) \\ &\underset{(*)}{\iff} \begin{pmatrix} \text{There exist } \mathbf{x}^1 \in F(\boldsymbol{\mathcal{C}}^1, \mu^{(1)}), \quad \mathbf{x}^2 \in F(\boldsymbol{\mathcal{C}}^2, \mu^{(2)}), \quad \mathbf{y}^1 \in \mathcal{X}^1, \text{ and } \mathbf{y}^2 \in \mathcal{X}^2 \\ \text{such that } \mathbf{d} = \mathbf{g}(\mathbf{y}, \mu) - \mathbf{g}(\mathbf{x}, \mu), \text{ where } \mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2) \text{ and } \mathbf{y} = (\mathbf{y}^1, \mathbf{y}^2) \end{pmatrix} \\ &\underset{(\dagger)}{\iff} \Big(\exists\, \mathbf{d}^1 \in \mathcal{D}_{\mathcal{H}}(F, \mathcal{X}^1, \mu^{(1)}) \text{ and } \mathbf{d}^2 \in \mathcal{D}_{\mathcal{H}}(F, \mathcal{X}^2, \mu^{(2)}) \text{ such that } \mathbf{d} = q_1\, p_1 \mathbf{d}^1 + q_2\, p_2 \mathbf{d}^2\Big). \end{aligned}$$

Here $(*)$ is because $\mathcal{X} = \mathcal{X}^1 \times \mathcal{X}^2$ and $F(\boldsymbol{\mathcal{C}}, \mu) = F(\boldsymbol{\mathcal{C}}^1, \mu^{(1)}) \times F(\boldsymbol{\mathcal{C}}^2, \mu^{(2)})$, because F is scale-invariant and satisfies Combination. (Furthermore, note that Scale invariance is only required if $p_1 \neq 1$ or $p_2 \neq 1$.) Meanwhile, $(\dagger)$ is by equations (C12) and (C14).

(d)
$$\begin{aligned} \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}) &\underset{(*)}{=} \bigcup_{\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})} \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}, \mu) \\ &\underset{(\dagger)}{=} \bigcup_{\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})} \big\{q_1\, p_1\, \mathbf{d}^1 + q_2\, p_2\, \mathbf{d}^2 \,;\; \mathbf{d}^1 \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^1, \mu^{(1)}) \text{ and } \mathbf{d}^2 \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^2, \mu^{(2)})\big\} \\ &\underset{(\diamond)}{=} \bigcup_{\mu_1 \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}}^1)} \bigcup_{\mu_2 \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}}^2)} \big\{q_1\, p_1\, \mathbf{d}^1 + q_2\, p_2\, \mathbf{d}^2 \,;\; \mathbf{d}^1 \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^1, \mu_1) \text{ and } \mathbf{d}^2 \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^2, \mu_2)\big\} \\ &\underset{(\ddagger)}{=} \big\{q_1\, p_1\, \mathbf{d}^1 + q_2\, p_2\, \mathbf{d}^2 \,;\; \mathbf{d}^1 \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^1) \text{ and } \mathbf{d}^2 \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^2\big\}. \end{aligned}$$

Here $(*)$ is by equation (C13), $(\dagger)$ is by part (c), $(\diamond)$ is by part (a), and $(\ddagger)$ is by two more applications of equation (C13). $\diamond$ `Claim 1`

Now we prove part (b) of the Lemma by induction on J. The case $J=2$ follows from Claim 1(d). So, let $J \geq 3$, and suppose the part (b) has been proved for the case $J-1$. Let $\boldsymbol{\mathcal{C}}' := p_1\, \boldsymbol{\mathcal{C}}^1 \times \cdots \times p_{J-1}\, \boldsymbol{\mathcal{C}}^{J-1}$, and for all $j \in [1 \ldots J-1]$, let $q'_j := |\boldsymbol{\mathcal{C}}_j|/|\boldsymbol{\mathcal{C}}'|$. By the induction hypothesis,

$$\mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}') = \left\{ q'_1\, p_1\, \mathbf{d}^1 + \cdots + q'_{J-1}\, p_{J-1}\, \mathbf{d}^{J-1} \ ; \ \mathbf{d}^j \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^j) \ \text{ for all } j \in [1 \ldots J-1] \right\}. \tag{C15}$$

Now let $\boldsymbol{\mathcal{C}} := p_1\, \boldsymbol{\mathcal{C}}^1 \times \cdots \times p_{J-1}\, \boldsymbol{\mathcal{C}}^{J-1} \times p_J\, \boldsymbol{\mathcal{C}}^J$. Let $q' := |\boldsymbol{\mathcal{C}}'|/|\boldsymbol{\mathcal{C}}|$ and let $q_J := |\boldsymbol{\mathcal{C}}^J|/|\boldsymbol{\mathcal{C}}|$. Note that $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{C}}' \times p_J\, \boldsymbol{\mathcal{C}}^J$. Thus,

$$\begin{aligned}
\mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}) \ &= \ \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}' \times p_J\, \boldsymbol{\mathcal{C}}^J) \underset{(*)}{=} \left\{ q'\, \mathbf{d}' + q_J\, p_J\, \mathbf{d}^J \ ; \ \mathbf{d}' \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}') \text{ and } \ \mathbf{d}^J \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^j \right\} \\
&\underset{(\dagger)}{=} \ \left\{ q'\, q'_1\, p_1\, \mathbf{d}^1 + \cdots + q'\, q'_{J-1}\, p_{J-1}\, \mathbf{d}^{J-1} + q_J\, p_J\, \mathbf{d}^J \ ; \ \mathbf{d}^j \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^j) \ \text{ for all } j \in [1 \ldots J] \right\}. \\
&\underset{(\diamond)}{=} \ \left\{ q_1\, p_1\, \mathbf{d}^1 + \cdots + q_{J-1}\, p_{J-1}\, \mathbf{d}^{J-1} + q_J\, p_J\, \mathbf{d}^J \ ; \ \mathbf{d}^j \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}^j) \ \text{ for all } j \in [1 \ldots J] \right\},
\end{aligned}$$

as claimed. Here, $(*)$ is by Claim 1(d). Next, $(\dagger)$ is by equation (C15). Finally, $(\diamond)$ is because $q'\, q'_j = q_j$ for all $j \in [1 \ldots J-1]$. $\square$

Proof of Lemma C.3. To show that $\mathcal{V}$ is finite, it suffices to show that $\mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}_j)$ is finite, for each $j \in [1 \ldots J]$. Suppose $\boldsymbol{\mathcal{C}}_j = (\mathcal{K}_j, \boldsymbol{\lambda}_j, \mathcal{X}_j, \mathcal{Y}_j)$. The set $\Delta_{\mathcal{H}}(\mathcal{Y}_j)$ is finite because $\mathcal{H}$ and $\mathcal{Y}_j$ are finite sets. Furthermore, for each $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y}_j)$ and $\mathbf{x} \in F(\boldsymbol{\mathcal{C}}_j, \mu)$, the set $\mathcal{D}_{\mathcal{H}}(\mathbf{x}, \boldsymbol{\mathcal{C}}_j, \mu)$ is finite, because $\mathcal{X}_j$ is finite. Thus, defining equation (C7) says that $\mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}_j)$ is a finite union of finite sets, hence finite. This proves part (a). The proofs of parts (b) and (c) involve four claims. Suppose F satisfies Combination on $\Delta_{\mathcal{H}}\langle\mathfrak{G}\rangle$.

Claim 1: $\operatorname{conv}_{\mathbb{Q}}(\mathcal{V}) \subseteq \mathcal{D}_{\mathcal{H}}(F, \langle\mathfrak{G}\rangle)$.

Proof: Let $N \in \mathbb{N}$, let $(q_1, \ldots, q_N) \in \Delta^N_{\mathbb{Q}}$, and let $\mathbf{v}_1, \ldots, \mathbf{v}_N \in \mathcal{V}$; we must show that $q_1\, \mathbf{v}_1 + \cdots + q_N\, \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \langle\mathfrak{G}\rangle)$.

Let M be a common denominator of $q_1, \ldots, q_N$, so that for all $n \in [1 \ldots N]$ there is some $m_n \in [0 \ldots M]$ such that $q_n = m_n/M$. (Thus, $m_1 + \cdots + m_N = M$.) For all $n \in [1 \ldots N]$, find some $j_n \in [1 \ldots J]$ such that $\mathbf{v}_n \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}_{j_n})$. Let $K_n := |\boldsymbol{\mathcal{C}}_{j_n}|$. Let $H := K_1\, K_2 \cdots K_N$, and for all $n \in [1 \ldots N]$, let $H_n := H/K_n$. Then H_n is an integer and $H_n\, K_n = H$. For all $n \in [1 \ldots N]$, let $L_n := m_n\, H_n$. Finally, let

$$K \ := \ L_1\, K_1 + \cdots + L_N K_N \ = \ \sum_{n=1}^{N} m_n\, H_n\, K_n \ = \ \sum_{n=1}^{N} m_n\, H \ = \ M\, H. \tag{C16}$$

Thus, for all $n \in [1 \ldots N]$,

$$\frac{L_n\, K_n}{K} \ = \ \frac{m_n\, H_n\, K_n}{M\, H} \ = \ \frac{m_n}{M} \ = \ q_n. \tag{C17}$$

Now define

$$\boldsymbol{\mathcal{C}} \ := \ \underbrace{\boldsymbol{\mathcal{C}}_{j_1} \times \cdots \times \boldsymbol{\mathcal{C}}_{j_1}}_{L_1} \times \underbrace{\boldsymbol{\mathcal{C}}_{j_2} \times \cdots \times \boldsymbol{\mathcal{C}}_{j_2}}_{L_2} \times \cdots \times \underbrace{\boldsymbol{\mathcal{C}}_{j_N} \times \cdots \times \boldsymbol{\mathcal{C}}_{j_N}}_{L_N}. \tag{C18}$$

Then $|\boldsymbol{\mathcal{C}}| = L_1 K_1 + \cdots + L_N K_N = K$. Now, for all $n \in [1 \dots N]$, let $\mathbf{d}_n^1, \dots, \mathbf{d}_n^{L_n} \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}_{j_n})$. Then combining formula (C18) with Lemma C.4(a), we obtain

$$\left(\sum_{\ell=1}^{L_1} \left(\frac{K_1}{K} \right) \mathbf{d}_1^{\ell} \right) + \left(\sum_{\ell=1}^{L_2} \left(\frac{K_2}{K} \right) \mathbf{d}_2^{\ell} \right) + \cdots + \left(\sum_{\ell=1}^{L_N} \left(\frac{K_N}{K} \right) \mathbf{d}_N^{\ell} \right) \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}). \quad \text{(C19)}$$

In particular, for all $n \in [1 \dots N]$, suppose that $\mathbf{d}_n^1 := \cdots := \mathbf{d}_n^{L_n} := \mathbf{v}_n$. Then formula (C19) becomes

$$\left(\frac{L_1 K_1}{K} \right) \mathbf{v}_1 + \cdots + \left(\frac{L_N K_N}{K} \right) \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}).$$

Applying (C17), this becomes $q_1 \mathbf{v}_1 + \cdots + q_N \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}})$. Thus, formula (C8) yields $q_1 \mathbf{v}_1 + \cdots + q_N \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G} \rangle)$, because $\boldsymbol{\mathcal{C}} \in \langle \mathfrak{G} \rangle$.

We can perform this construction for any $(q_1, \dots, q_N) \in \Delta_{\mathbb{Q}}^N$ and $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathcal{V}$. Thus, we conclude that $\mathrm{conv}_{\mathbb{Q}}(\mathcal{V}) \subseteq \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G} \rangle)$, as desired. $\diamond$ Claim 1

Claim 2: $\mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G} \rangle) \subseteq \mathrm{conv}_{\mathbb{Q}}(\mathcal{V})$.

Proof: For any $j_1, \dots, j_N \in [1 \dots J]$, Lemma C.4(a) implies that $\mathcal{D}_{\mathcal{H}}(F,\ \boldsymbol{\mathcal{C}}_{j_1} \times \cdots \times \boldsymbol{\mathcal{C}}_{j_N}) \subseteq \mathrm{conv}_{\mathbb{Q}}(\mathcal{V})$. Thus,

$$\mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G} \rangle) \underset{(*)}{=} \bigcup_{\boldsymbol{\mathcal{C}} \in \langle \mathfrak{G} \rangle} \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}) \underset{(\dagger)}{=} \bigcup_{N=1}^{\infty} \bigcup_{j_1, \dots, j_N = 1}^{J} \mathcal{D}_{\mathcal{H}}(F,\ \boldsymbol{\mathcal{C}}_{j_1} \times \cdots \times \boldsymbol{\mathcal{C}}_{j_N}) \subseteq \mathrm{conv}_{\mathbb{Q}}(\mathcal{V}),$$

as desired. Here $(*)$ is by formula (C8), and $(\dagger)$ is by formula (5.1). $\diamond$ Claim 2

Claims 1 and 2 together yield $\mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G} \rangle) = \mathrm{conv}_{\mathbb{Q}}(\mathcal{V})$; this proves part (b).

To prove part (c), suppose F satisfies Combination and Scale invariance on $\Delta_{\mathcal{H}} \langle \mathfrak{G}, \epsilon \rangle$.

Claim 3: $\mathrm{conv}(\mathcal{V}) \subseteq \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G}, \epsilon \rangle)$.

Proof: Let $N \in \mathbb{N}$, let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathcal{V}$ and let $(r_1, \dots, r_N) \in \Delta^N$. We must show that $\sum_{n=1}^{N} r_n \mathbf{v}_n \in \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G}, \epsilon \rangle)$. First, note that we can assume without loss of generality that $r_n > 0$ for all $n \in [1 \dots N]$. Let $r_* := \min\{r_1, \dots, r_N\}$; then $r_* > 0$.

Now, $\Delta_{\mathbb{Q}}^N$ is a dense subset of Δ^N. Thus, there exists some $(q_1, \dots, q_N) \in \Delta_{\mathbb{Q}}^N$ such that $q_n > 0$ and $|q_n - r_n| < \frac{r_* \epsilon}{2}$ for all $n \in [1 \dots N]$. For each $n \in [1 \dots N]$, let $p_n := r_n / q_n$. Then $p_n \in \mathbb{R}_+$ (because $r_n \in \mathbb{R}_+$ and $q_n \in \mathbb{Q}_+ \subseteq \mathbb{R}_+$), and it is easily verified that

$$1 - \epsilon \quad < \quad p_n \quad < \quad 1 + \epsilon, \quad \text{for all } n \in [1 \dots N]. \quad \text{(C20)}$$

(Proof sketch: For any $n \in [1 \dots N]$, $-\frac{r_* \epsilon}{2} < r_n - q_n < \frac{r_* \epsilon}{2}$, and thus, $1 - \frac{r_* \epsilon}{2 q_n} < p_n < 1 + \frac{r_* \epsilon}{2 q_n}$. But $q_n > r_n - \frac{r_* \epsilon}{2} > r_n - \frac{r_*}{2} \geq r_* - \frac{r_*}{2} = \frac{r_*}{2}$. Thus, $\frac{r_* \epsilon}{2 q_n} < \epsilon$.) Meanwhile,

$$q_1 p_1 + \cdots + q_N p_N \quad = \quad r_1 + \cdots + r_N \quad = \quad 1. \quad \text{(C21)}$$

For all $n \in [1 \ldots N]$, find $j_n \in [1 \ldots J]$ such that $\mathbf{v}_n \in \boldsymbol{\mathcal{C}}_{j_n}$. Let $K_n := |\boldsymbol{\mathcal{C}}_{j_n}|$. Now define M, $m_1, \ldots, m_N$, H, $H_1, \ldots, H_N$, and $L_1, \ldots, L_N$ as in the proof of Claim 1, so that equations (C16) and (C17) are satisfied. Thus,

$$\sum_{\ell=1}^{L_1} p_1\, K_1 + \cdots + \sum_{\ell=1}^{L_N} p_N\, K_N \;=\; p_1\, L_1\, K_1 + \cdots + p_N\, L_N\, K_N \underset{(*)}{=} p_1\, q_1\, K + \cdots + p_N\, q_N\, K \underset{(\dagger)}{=} K,$$

where $(*)$ is because equation (C17) implies that $L_n\, K_n = q_n\, K$ for all $n \in [1 \ldots N]$, and $(\dagger)$ is by equation (C21). Combining this observation with the inequalities (C20), and comparing to the defining formula (C10), we deduce that

$$(\underbrace{p_1, \ldots, p_1}_{L_1}, \underbrace{p_2, \ldots, p_2}_{L_2}, \ldots, \underbrace{p_N, \ldots, p_N}_{L_N}) \;\in\; \mathcal{P}_\epsilon(\underbrace{K_1, \ldots, K_1}_{L_1}, \underbrace{K_2, \ldots, K_2}_{L_2}, \ldots, \underbrace{K_N, \ldots, K_N}_{L_N}).$$

Thus, if we define

$$\boldsymbol{\mathcal{C}} \;:=\; \underbrace{p_1\, \boldsymbol{\mathcal{C}}_{j_1} \times \cdots \times p_1\, \boldsymbol{\mathcal{C}}_{j_1}}_{L_1} \times \underbrace{p_2\, \boldsymbol{\mathcal{C}}_{j_2} \times \cdots \times p_2\, \boldsymbol{\mathcal{C}}_{j_2}}_{L_2} \times \cdots \times \underbrace{p_N\, \boldsymbol{\mathcal{C}}_{j_N} \times \cdots \times p_N\, \boldsymbol{\mathcal{C}}_{j_N}}_{L_N}, \tag{C22}$$

then $\boldsymbol{\mathcal{C}} \in \langle \mathfrak{G}, \epsilon \rangle$, by defining formula (C11).

Now, for all $n \in [1 \ldots N]$, let $\mathbf{d}_n^1, \ldots, \mathbf{d}_n^{L_n} \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}_{j_n})$. Then combining formula (C18) with Lemma C.4(b), we obtain

$$\left(\sum_{\ell=1}^{L_1} \left(\frac{K_1}{K} \right) p_1\, \mathbf{d}_1^\ell \right) + \left(\sum_{\ell=1}^{L_2} \left(\frac{K_2}{K} \right) p_2\, \mathbf{d}_2^\ell \right) + \cdots + \left(\sum_{\ell=1}^{L_N} \left(\frac{K_N}{K} \right) p_N\, \mathbf{d}_N^\ell \right) \;\in\; \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}). \tag{C23}$$

In particular, for all $n \in [1 \ldots N]$, suppose that $\mathbf{d}_n^1 := \cdots := \mathbf{d}_n^{L_n} := \mathbf{v}_n$. Then formula (C23) becomes

$$\left(\frac{p_1\, L_1\, K_1}{K} \right) \mathbf{v}_1 + \cdots + \left(\frac{p_N\, L_N\, K_N}{K} \right) \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}}).$$

Applying (C17), this becomes $p_1\, q_1\, \mathbf{v}_1 + \cdots + p_N\, q_N\, \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}})$. But $p_n\, q_n = r_n$ for all $n \in [1 \ldots N]$. Thus, we get $r_1\, \mathbf{v}_1 + \cdots + r_N\, \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \boldsymbol{\mathcal{C}})$. Thus, formula (C8) yields $r_1\, \mathbf{v}_1 + \cdots + r_N\, \mathbf{v}_N \in \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G}, \epsilon \rangle)$, because $\boldsymbol{\mathcal{C}} \in \langle \mathfrak{G}, \epsilon \rangle$.

We can perform this construction for any $(r_1, \ldots, r_N) \in \Delta^N$ and $\mathbf{v}_1, \ldots, \mathbf{v}_N \in \mathcal{V}$. Thus, we conclude that $\mathrm{conv}(\mathcal{V}) \subseteq \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G}, \epsilon \rangle)$, as desired. ◇ `Claim 3`

Claim 4: $\mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G}, \epsilon \rangle) \subseteq \mathrm{conv}(\mathcal{V})$.

Proof: $\mathcal{D}_{\mathcal{H}}(F,\langle\mathfrak{G},\epsilon\rangle) \underset{(*)}{=} \bigcup_{\boldsymbol{\mathcal{C}}\in\langle\mathfrak{G},\epsilon\rangle} \mathcal{D}_{\mathcal{H}}(F,\boldsymbol{\mathcal{C}})$

$$
\begin{aligned}
&\underset{(\dagger)}{=} \bigcup_{N=1}^{\infty}\ \bigcup_{j_1,\ldots,j_N=1}^{J}\ \bigcup_{\mathbf{p}\in\mathcal{P}_\epsilon(|\boldsymbol{\mathcal{C}}_{j_1}|,\ldots,|\boldsymbol{\mathcal{C}}_{j_N}|)} \mathcal{D}_{\mathcal{H}}(F,\ p_1\,\boldsymbol{\mathcal{C}}_{j_1}\times\cdots\times p_N\,\boldsymbol{\mathcal{C}}_{j_N})\\
&\underset{(\diamond)}{=} \bigcup_{N=1}^{\infty}\ \bigcup_{j_1,\ldots,j_N=1}^{J}\ \bigcup_{\mathbf{p}\in\mathcal{P}_\epsilon(|\boldsymbol{\mathcal{C}}_{j_1}|,\ldots,|\boldsymbol{\mathcal{C}}_{j_N}|)} \{q_1\,p_1\,\mathbf{d}_1+\cdots+q_N\,p_N\,\mathbf{d}_N\ ;\ \mathbf{d}_n\in\mathcal{D}_{\mathcal{H}}(F,\boldsymbol{\mathcal{C}}_{j_N}),\ \forall\ n\in[1...N]\}\\
&\underset{(\ddagger)}{\subseteq} \bigcup_{N=1}^{\infty}\ \bigcup_{j_1,\ldots,j_N=1}^{J}\ \bigcup_{\mathbf{r}\in\Delta^N} \{r_1\,\mathbf{d}_1+\cdots+r_N\,\mathbf{d}_N\ ;\ \mathbf{d}_n\in\mathcal{D}_{\mathcal{H}}(F,\boldsymbol{\mathcal{C}}_{j_N}),\ \forall\ n\in[1...N]\}\\
&\underset{(@)}{=} \left\{r_1\,\mathbf{v}_1+\cdots+r_N\,\mathbf{v}_N\ ;\ N\in\mathbb{N},\ \mathbf{v}_N\in\mathcal{V}\text{ for all }n\in[1\ldots N]\text{ and }\ \mathbf{r}\in\Delta^J\right\}\\
&= \operatorname{conv}(\mathcal{V}),
\end{aligned}
$$

as claimed. Here, $(*)$ is by defining equation (C8) and $(\dagger)$ is by defining equation (C11). Meanwhile, $(\diamond)$ is by Lemma C.4(b), with the convention that $q_m := |\boldsymbol{\mathcal{C}}_{j_m}|/\sum_{n=1}^{N}|\boldsymbol{\mathcal{C}}_{j_n}|$, for all $m\in[1\ldots N]$. Next, $(\ddagger)$ is because defining formula (C10) implies that $(p_1\,q_1,\ldots,p_N\,q_N)\in\Delta^N$ for any $(p_1,\ldots,p_N)\in\mathcal{P}_\epsilon(|\boldsymbol{\mathcal{C}}_{j_1}|,\ldots,|\boldsymbol{\mathcal{C}}_{j_N}|)$. Finally, $(@)$ is by the definition of $\mathcal{V}$. $\diamondsuit$ `Claim 4`

Claims 3 and 4 imply that $\mathcal{D}_{\mathcal{H}}(F,\langle\mathfrak{G},\epsilon\rangle)=\operatorname{conv}(\mathcal{V})$; this proves part (c). $\Box$

The next lemma is actually Theorem 1(a).

Lemma C.5 *Any additive majority rule satisfies* Supermajority efficiency, Combination, *and* Scale invariance *in any judgement environment.*

Proof: Let $\phi:[-1,1]\longrightarrow{}^*\mathbb{R}$ be a gain function.

Supermajority efficiency. (by contrapositive) Let $\boldsymbol{\mathcal{C}}=(\mathcal{K},\boldsymbol{\lambda},\mathcal{X},\mathcal{Y})$ be a judgement context. Let $\mu\in\Delta(\mathcal{Y})$. We must show that $F_\phi(\boldsymbol{\mathcal{C}};\mu)\subseteq\operatorname{SME}(\boldsymbol{\mathcal{C}},\mu)$. Contrapositively, let $\mathbf{y}\in\mathcal{X}$, and suppose $\mathbf{y}\notin\operatorname{SME}(\boldsymbol{\mathcal{C}},\mu)$; we will show that $\mathbf{y}\notin F_\phi(\boldsymbol{\mathcal{C}};\mu)$.

Let $\mathcal{H}_\mu := \{|\widetilde{\mu}_k|;\ \ k\in\mathcal{K}\}$. For all $\mathbf{x}\in\mathcal{X}$, and all $h\in\mathcal{H}_\mu$, let $\Lambda_h(\mathbf{x}) := \sum\{\lambda_k;\ k\in\mathcal{M}(\mathbf{x},\mu)\text{ and }|\widetilde{\mu}_k|=h\}$. Then Proposition 4.3(a) implies that

$$F_\phi(\boldsymbol{\mathcal{C}};\mu) \quad=\quad \operatorname*{argmax}_{\mathbf{x}\in\mathcal{X}} \sum_{h\in\mathcal{H}_\mu}\Lambda_h(\mathbf{x})\,\phi(h). \tag{C24}$$

Also note that, for all $r\in[0,1]$, we have

$$\gamma^{\boldsymbol{\lambda}}_{\mathbf{x},\mu}(r) \quad=\quad \sum_{h\in\mathcal{H}_\mu;\ h\geq r}\Lambda_h(\mathbf{x}). \tag{C25}$$

Now, if $\mathbf{y}\notin\operatorname{SME}(\boldsymbol{\mathcal{C}},\mu)$, then there is some $\mathbf{x}\in\mathcal{X}$ such that $\gamma^{\boldsymbol{\lambda}}_{\mathbf{x},\mu}(r)\geq\gamma^{\boldsymbol{\lambda}}_{\mathbf{y},\mu}(r)$ for all for all $r\in[0,1]$, with strict inequality for some $r\in[0,1]$. From formula (C25), we see that this means that $\Lambda_h(\mathbf{x}) \geq \Lambda_h(\mathbf{y})$ for all $h \in \mathcal{H}_\mu$, with strict inequality for some $h \in \mathcal{H}_\mu$. But then $\sum_{h\in\mathcal{H}_\mu} \Lambda_h(\mathbf{x})\,\phi(h) > \sum_{h\in\mathcal{H}_\mu} \Lambda_h(\mathbf{y})\,\phi(h)$. Thus, formula (C24) implies that $\mathbf{y} \notin F_\phi(\boldsymbol{\mathcal{C}}, \mu)$.

Combination and Scale invariance. Scale invariance is easily verified, so we focus on Combination. Let $\mathfrak{E}$ be a judgement environment. It suffices to prove the case $J = 2$ of the Combination axiom. (The general case then follows by induction on J.) So, let $\boldsymbol{\mathcal{C}}_1 = (\mathcal{K}_1, \boldsymbol{\lambda}_1, \mathcal{X}_1, \mathcal{Y}_1)$ and $\boldsymbol{\mathcal{C}}_2 = (\mathcal{K}_2, \boldsymbol{\lambda}_2, \mathcal{X}_2, \mathcal{Y}_2)$ be two judgement contexts in $\mathfrak{E}$, and let $(\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y}) = \boldsymbol{\mathcal{C}}_1 \times \boldsymbol{\mathcal{C}}_2$. Thus, $\mathcal{K} = \mathcal{K}_1 \sqcup \mathcal{K}_2$, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$, and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$. Let $\mu \in \Delta(\mathcal{Y})$, and let $\mu^{(1)}$ and $\mu^{(2)}$ be the marginal projections of μ onto $\Delta(\mathcal{Y}_1)$ and $\Delta(\mathcal{Y}_2)$ respectively, as defined by formula (5.3).

We claim that $F_\phi(\boldsymbol{\mathcal{C}}; \mu) = F_\phi(\boldsymbol{\mathcal{C}}_1; \mu^{(1)}) \times F_\phi(\boldsymbol{\mathcal{C}}_2; \mu^{(2)})$. If $\mathbf{x} \in \mathcal{X}$, then $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ for some $\mathbf{x}_1 \in \mathcal{X}_1$ and $\mathbf{x}_2 \in \mathcal{X}_2$. Meanwhile, $\widetilde{\boldsymbol{\mu}} = (\widetilde{\boldsymbol{\mu}}^{(1)}, \widetilde{\boldsymbol{\mu}}^{(2)})$, so that $\phi(\widetilde{\boldsymbol{\mu}}) = (\phi(\widetilde{\boldsymbol{\mu}}^{(1)}), \phi(\widetilde{\boldsymbol{\mu}}^{(2)}))$, while $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$. Thus, $\phi(\widetilde{\boldsymbol{\mu}}) \underset{\boldsymbol{\lambda}}{\bullet} \mathbf{x} = \phi(\widetilde{\boldsymbol{\mu}}^{(1)}) \underset{\boldsymbol{\lambda}_1}{\bullet} \mathbf{x}_1 + \phi(\widetilde{\boldsymbol{\mu}}^{(2)}) \underset{\boldsymbol{\lambda}_2}{\bullet} \mathbf{x}_2$. Thus,

$$\begin{aligned}
&\Big(\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}}; \mu)\Big) \iff \Big(\phi(\widetilde{\boldsymbol{\mu}}) \underset{\boldsymbol{\lambda}}{\bullet} \mathbf{x} \geq \phi(\widetilde{\boldsymbol{\mu}}) \underset{\boldsymbol{\lambda}}{\bullet} \mathbf{y} \text{ for all } \mathbf{y} \in \mathcal{X}\Big) \\
&\iff \Big(\phi(\widetilde{\boldsymbol{\mu}}^{(1)}) \underset{\boldsymbol{\lambda}_1}{\bullet} \mathbf{x}_1 \geq \phi(\widetilde{\boldsymbol{\mu}}^{(1)}) \underset{\boldsymbol{\lambda}_1}{\bullet} \mathbf{y}_1,\ \forall\, \mathbf{y}_1 \in \mathcal{X}_1 \text{ and } \phi(\widetilde{\boldsymbol{\mu}}^{(2)}) \underset{\boldsymbol{\lambda}_2}{\bullet} \mathbf{x}_2 \geq \phi(\widetilde{\boldsymbol{\mu}}^{(2)}) \underset{\boldsymbol{\lambda}_2}{\bullet} \mathbf{y}_2,\ \forall\, \mathbf{y}_2 \in \mathcal{X}_2\Big) \\
&\iff \Big(\mathbf{x}_1 \in F_\phi(\boldsymbol{\mathcal{C}}_1; \mu^{(1)}) \text{ and } \mathbf{x}_2 \in F_\phi(\boldsymbol{\mathcal{C}}_2; \mu^{(2)})\Big) \iff \Big(\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}}_1; \mu^{(1)}) \times F_\phi(\boldsymbol{\mathcal{C}}_2; \mu^{(2)})\Big),
\end{aligned}$$

as desired. □

Lemma C.6 *Let $\mathfrak{G}$ be a set of judgement contexts, and let F_1 and F_2 be two judgement aggregation rules such that $F_1(\boldsymbol{\mathcal{G}}, \mu) \subseteq F_2(\boldsymbol{\mathcal{G}}, \mu)$ for all $\boldsymbol{\mathcal{G}} \in \mathfrak{G}$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{G}})$.*

- **(a)** *If F_1 and F_2 satisfy* Combination *on $\Delta_{\mathcal{H}}\langle\mathfrak{G}\rangle$, then $F_1(\boldsymbol{\mathcal{C}}, \mu) \subseteq F_2(\boldsymbol{\mathcal{C}}, \mu)$ for all $\boldsymbol{\mathcal{C}} \in \langle\mathfrak{G}\rangle$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$.*

- **(b)** *Suppose $\mathfrak{E}$ is a judgement environment that satisfies condition* (FG2) *with respect to $\mathfrak{G}$. If F_1 and F_2 satisfy* Combination *and* Scale invariance *on $\Delta_{\mathcal{H}}(\mathfrak{E})$, then $F_1(\boldsymbol{\mathcal{C}}, \mu) \subseteq F_2(\boldsymbol{\mathcal{C}}, \mu)$ for all $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$.*

Proof: We will prove part (b); the proof of (a) is very similar (but does not require Scale invariance). Suppose $\mathfrak{G} = \{\boldsymbol{\mathcal{G}}_1, \ldots, \boldsymbol{\mathcal{G}}_N\}$. By (FG2), any $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$ has the form $\boldsymbol{\mathcal{C}} = p_1\boldsymbol{\mathcal{C}}_{n_1} \times \cdots \times p_J\boldsymbol{\mathcal{C}}_{n_J}$ for some $p_1, \ldots, p_J \in \mathbb{R}_+$ and $n_1, \ldots, n_J \in [1 \ldots N]$. Suppose $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ (where $\mathcal{Y} = \mathcal{Y}_{n_1} \times \cdots \times \mathcal{Y}_{n_J}$, etc.) For any $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$ and any $j \in [1 \ldots J]$, let $\mu^{(j)} \in \Delta_{\mathcal{H}}(\mathcal{Y}_{n_j})$ be the nth marginal of μ, as defined by formula (5.3). Since F_1 and F_2 satisfy Combination and Scale invariance, we have:

$$F_1(\boldsymbol{\mathcal{C}}; \mu) \;=\; \prod_{j=1}^{J} F_1(\boldsymbol{\mathcal{G}}_{n_j}; \mu^{(j)}) \quad \text{and} \quad F_2(\boldsymbol{\mathcal{C}}; \mu) \;=\; \prod_{j=1}^{J} F_2(\boldsymbol{\mathcal{G}}_{n_j}; \mu^{(j)}).$$

By hypothesis, $F_1(\boldsymbol{\mathcal{G}}_{n_j}; \mu^{(j)}) \subseteq F_2(\boldsymbol{\mathcal{G}}_{n_j}; \mu^{(j)})$ for all $j \in [1 \ldots J]$. Thus, $F_1(\boldsymbol{\mathcal{C}}, \mu) \subseteq F_2(\boldsymbol{\mathcal{C}}, \mu)$, as desired. □

Lemma C.7 *Let $\psi : [-1,1]{\longrightarrow}{}^*\mathbb{R}$ be a gain function. Suppose that either*

(a) *$\mathfrak{E} = \langle \mathfrak{G} \rangle$ for some finite collection $\mathfrak{G}$ of judgement contexts; or*

(b) *$\mathfrak{E}$ is a finitely generated judgement environment.*

For any $\mathcal{H} \in \wp[0,1]$, there exists a real-valued gain function $\phi_{\mathcal{H}} : \mathcal{H}{\longrightarrow}\mathbb{R}$ such that $F_{\phi_{\mathcal{H}}}(\boldsymbol{\mathcal{C}};\mu) = F_\psi(\boldsymbol{\mathcal{C}};\mu)$ for all $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$ and all $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$.

Proof: We will prove Case (b) of the lemma; the proof of Case (a) is similar. Suppose $\mathfrak{E}$ is generated by the set $\mathfrak{G} = \{\boldsymbol{\mathcal{C}}_1, \ldots, \boldsymbol{\mathcal{C}}_N\}$. Lemma C.5 says that F_ψ and $F_{\phi_{\mathcal{H}}}$ satisfy Combination and Scale invariance on $\Delta_{\mathcal{H}}(\mathfrak{E})$. Thus, Lemma C.6(b) tells us that it suffices to construct $\phi_{\mathcal{H}}$ such that $F_{\phi_{\mathcal{H}}}(\boldsymbol{\mathcal{C}}_n;\mu) = F_\psi(\boldsymbol{\mathcal{C}}_n;\mu)$ for all $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}}_n)$ and all $n \in [1 \ldots N]$.

Now, ${}^*\mathbb{R}$ is a linearly ordered vector space. Thus, Hahn's Embedding Theorem says there is an order-preserving linear isomorphism $\alpha : {}^*\mathbb{R}{\longrightarrow}\mathcal{L} \subset \mathbb{R}^{\mathcal{T}}$, where $\mathcal{T}$ is a (possibly infinite) linearly ordered set, and the linear subspace $\mathcal{L}$ is endowed with the $\mathcal{T}$-lexicographical order, with smaller t-coordinates given lexicographical priority over larger t coordinates; see e.g. Hausner and Wendel (1952), Clifford (1954), or Gravett (1956) for details. (In particular, this means that, for any distinct $\ell, \ell' \in \mathcal{L}$, the set $\{t \in \mathcal{T};\ \ell_t \neq \ell'_t\}$ has a minimal element.)[30]

For all $n \in [1 \ldots N]$, suppose that $\boldsymbol{\mathcal{C}}_n = (\mathcal{K}_n, \boldsymbol{\lambda}^n, \mathcal{X}_n, \mathcal{Y}_n)$. Now, let $n \in [1 \ldots N]$. For any $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y}_n)$, and any $\mathbf{x}, \mathbf{y} \in \mathcal{X}_n$, if $\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) \neq \mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})$, then $\alpha\left[\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right] \neq \alpha\left[\mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right]$ (because α is injective), and thus, there exists some $t^\mu_{\mathbf{x},\mathbf{y}} \in \mathcal{T}$ such that $\alpha\left[\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right]_{t^\mu_{\mathbf{x},\mathbf{y}}} \neq \alpha\left[\mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right]_{t^\mu_{\mathbf{x},\mathbf{y}}}$, while $\alpha\left[\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right]_t = \alpha\left[\mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right]_t$ for all $t < t^\mu_{\mathbf{x},\mathbf{y}}$. Since α is order-preserving and $\mathcal{L}$ is lexicographically ordered, this implies that

$$\left(\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) > \mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right) \quad \Longleftrightarrow \quad \left(\alpha\left[\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right]_{t^\mu_{\mathbf{x},\mathbf{y}}} > \alpha\left[\mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}})\right]_{t^\mu_{\mathbf{x},\mathbf{y}}}\right). \quad \text{(C26)}$$

Let $\mathcal{T}' := \{t^\mu_{\mathbf{x},\mathbf{y}};\ \ n \in [1 \ldots N],\ \ \mu \in \Delta_{\mathcal{H}}(\mathcal{Y}_n), \text{ and } \mathbf{x}, \mathbf{y} \in \mathcal{X}_n\}$. Then $\mathcal{T}'$ is finite, because $\mathcal{X}_n$ and $\Delta_{\mathcal{H}}(\mathcal{Y}_n)$ are finite for all $n \in [1 \ldots N]$. Let $<'$ denote the linear order which $\mathcal{T}'$ inherits from $\mathcal{T}$. For any $t \in \mathcal{T}'$, let $|t| := \#\{t' \in \mathcal{T}';\ \ t' <' t\}$; thus, $|t| \in \mathbb{N}$, because $\mathcal{T}'$ is finite. For any $t_1, t_2 \in \mathcal{T}'$, clearly $(t_1 < t_2) \Leftrightarrow (|t_1| < |t_2|)$.

Since $\mathcal{L} \subset \mathbb{R}^{\mathcal{T}}$, any $\ell \in \mathcal{L}$ has the form $\ell = (\ell_t)_{t \in \mathcal{T}}$ where $\ell_t \in \mathbb{R}$ for all $t \in \mathcal{T}$. Thus, for any $\epsilon > 0$, we can define a linear function $\beta_\epsilon : \mathcal{L}{\longrightarrow}\mathbb{R}$ by setting

$$\beta_\epsilon(\ell) \quad := \quad \sum_{t \in \mathcal{T}'} \ell_t \, \epsilon^{|t|}, \qquad \text{for all } \ell \in \mathcal{L}. \qquad \text{(C27)}$$

Then define $\phi_\epsilon : \mathcal{H}{\longrightarrow}\mathbb{R}$ by setting

$$\phi_\epsilon(h) \quad := \quad \beta_\epsilon\left(\alpha\left[\psi(h)\right]\right) \quad = \quad \sum_{t \in \mathcal{T}'} \alpha\left[\psi(h)\right]_t \, \epsilon^{|t|}, \qquad \text{for all } h \in \mathcal{H}. \qquad \text{(C28)}$$

[30]This technical remark is necessary because while $\mathcal{T}$ is linearly ordered, it may not be *well*-ordered, so the "lexicographical order" is not necessarily well-defined on all of $\mathbb{R}^{\mathcal{T}}$. But Hahn's Embedding Theorem ensures that it *is* well-defined on the subspace $\mathcal{L}$.

Then, for any $n \in [1 \ldots N]$, any $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y}_n)$, and $\mathbf{x}, \mathbf{y} \in \mathcal{X}_n$, we have

$$\begin{aligned}
\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \phi_\epsilon(\widetilde{\boldsymbol{\mu}}) - \mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \phi_\epsilon(\widetilde{\boldsymbol{\mu}}) \quad &\underset{(*)}{=} \quad \sum_{k \in \mathcal{K}_n} \lambda_k^n \, x_k \, \beta_\epsilon \left(\alpha \left[\psi(\widetilde{\mu}_k) \right] \right) - \sum_{k \in \mathcal{K}_n} \lambda_k^n \, y_k \, \beta_\epsilon \left(\alpha \left[\psi(\widetilde{\mu}_k) \right] \right) \\
\underset{(\dagger)}{=} \quad & \beta_\epsilon \left(\alpha \left[\sum_{k \in \mathcal{K}_n} \lambda_k^n \, x_k \, \psi(\widetilde{\mu}_k) \right] \right) - \beta_\epsilon \left(\alpha \left[\sum_{k \in \mathcal{K}_n} \lambda_k^n \, y_k \, \psi(\widetilde{\mu}_k) \right] \right) \\
= \quad & \beta_\epsilon \left(\alpha \left[\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) \right] \right) - \beta_\epsilon \left(\alpha \left[\mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) \right] \right) \\
\underset{(\diamond)}{=} \quad & \left(\alpha \left[\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) \right]_{t^\mu_{\mathbf{x},\mathbf{y}}} - \alpha \left[\mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) \right]_{t^\mu_{\mathbf{x},\mathbf{y}}} \right) \epsilon^{|t^\mu_{\mathbf{x},\mathbf{y}}|} \; + \; \begin{pmatrix} \text{a finite linear combination} \\ \text{of higher powers of } \epsilon \end{pmatrix}.
\end{aligned}$$

(Here, $(*)$ is by equation (C28) and the definition of $\mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi_\epsilon(\widetilde{\boldsymbol{\mu}})$. Next, $(\dagger)$ is because β_ϵ and α are both linear functions, and $x_k, y_k, \lambda_k^n \in \mathbb{R}$ for all $k \in \mathcal{K}_n$. Finally, $(\diamond)$ is by equation (C27) and the definition of $t^\mu_{\mathbf{x},\mathbf{y}}$.) Thus, there exists some $\epsilon^\mu_{\mathbf{x},\mathbf{y}} > 0$ such that, for any $\epsilon \in (0, \epsilon^\mu_{\mathbf{x},\mathbf{y}}]$, statement (C26) entails

$$\left(\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) > \mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \psi(\widetilde{\boldsymbol{\mu}}) \right) \iff \left(\mathbf{x} \underset{\boldsymbol{\lambda}^n}{\bullet} \phi_\epsilon(\widetilde{\boldsymbol{\mu}}) > \mathbf{y} \underset{\boldsymbol{\lambda}^n}{\bullet} \phi_\epsilon(\widetilde{\boldsymbol{\mu}}) \right). \tag{C29}$$

Now define $\epsilon := \min\{\epsilon^\mu_{\mathbf{x},\mathbf{y}};\ \ \mathbf{x}, \mathbf{y} \in \mathcal{X}_n,\ \mu \in \Delta_{\mathcal{H}}(\mathcal{Y}_n), \text{ and } n \in [1 \ldots N]\}$. Then $\epsilon > 0$ because $\mathcal{X}_n$ and $\Delta_{\mathcal{H}}(\mathcal{Y}_n)$ are finite for all $n \in [1 \ldots N]$. Furthermore, for any $n \in [1 \ldots N]$, and any $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y}_n)$, statement (C29) now holds for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_n$. Thus, Proposition 4.3(b) implies that

$$F_{\phi_\epsilon}(\boldsymbol{\mathcal{C}}_n; \mu) = F_\psi(\boldsymbol{\mathcal{C}}_n; \mu), \quad \text{for all } n \in [1 \ldots N] \text{ and } \mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}}_n). \tag{C30}$$

Now define $\phi_{\mathcal{H}} := \phi_\epsilon$. Then Lemma C.6 and equation (C30) imply that $F_{\phi_{\mathcal{H}}}(\boldsymbol{\mathcal{C}}; \mu) = F_\psi(\boldsymbol{\mathcal{C}}; \mu)$ for all $\boldsymbol{\mathcal{C}} \in \langle \mathfrak{G} \rangle$ and all $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$, as desired. □

Finally, we come to the proofs of the main results of this appendix.

Proof of Theorem C.2. (a) "$\Longleftarrow$" If $\phi_{\mathcal{H}}$ is strictly increasing, then for all $\boldsymbol{\mathcal{C}} \in \mathfrak{C}$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$, Lemma C.5 says $F_{\phi_{\mathcal{H}}}(\mu) \subseteq \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$, and thus, $F(\mu) \subseteq \mathrm{SME}\,(\boldsymbol{\mathcal{C}}, \mu)$. Thus, F satisfies Supermajority efficiency on $\Delta_{\mathcal{H}}(\mathfrak{C})$.

"$\Longrightarrow$" Suppose $\mathfrak{C}$ is finitely generated by a finite collection $\mathfrak{G}$ of judgement contexts, and let $\epsilon > 0$ be as in condition (FG3). By hypothesis, F is satisfies Supermajority efficiency, Combination and Scale invariance on $\Delta_{\mathcal{H}}(\mathfrak{C})$; thus, Lemmas C.1, C.2 and C.3(a,c) together yield an all-positive vector $\mathbf{z} \in \mathbb{R}^{\mathcal{H}}_{++}$ such that $\mathbf{z} \bullet \mathbf{p} \leq 0$ for all $\mathbf{p} \in \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G}, \epsilon \rangle)$. Now let $\pm\mathcal{H} := \{\pm h;\ h \in \mathcal{H}\}$, and define $\phi : \pm\mathcal{H} \longrightarrow \mathbb{R}$ as follows: for all $r \in \pm\mathcal{H}$, let

$$\phi(r) \quad := \quad \sum_{\substack{h \in \mathcal{H} \\ h \leq r}} z_h \ \text{ if } r \geq 0, \text{ and } \ \phi(r) \quad := \quad -\phi(-r) \ \text{ if } r \leq 0. \tag{C31}$$

Then ϕ is odd. Also, ϕ is increasing on $\pm\mathcal{H}$, because $z_h > 0$ for all $h \in \mathcal{H}$.

Let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y}) \in \langle \mathfrak{G}, \epsilon \rangle$ and let $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$. For any $\mathbf{x} \in \mathcal{X}$, recall $\mathcal{M}(\mu, \mathbf{x}) := \{k \in \mathcal{K};\ \ x_k \, \widetilde{\mu}_k \geq 0\}$.

Claim 1: *For any* $\mathbf{x} \in \mathcal{X}$, *we have* $\mathbf{z} \bullet \mathbf{g}(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu) = \displaystyle\sum_{k \in \mathcal{M}(\mathbf{x}, \mu)} \lambda_k \, \phi \, |\widetilde{\mu}_k|$.

Proof: For any $g, h \in [-1, 1]$, define $\delta_h^g := 1$ if $g \geq h$, whereas $\delta_h^g := 0$ if $g < h$. Let $K := |\mathcal{K}|$. Then

$$\begin{aligned}
\mathbf{z} \bullet \mathbf{g}(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu) &= \sum_{h \in \mathcal{H}} z_h \cdot g_h(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu) \underset{(\diamond)}{=} \sum_{h \in \mathcal{H}} z_h \cdot \frac{1}{K} \sum \{\lambda_k \; ; \; k \in \mathcal{K} \text{ and } x_k \, \widetilde{\mu}_k \geq h\} \\
&= \frac{1}{K} \sum_{h \in \mathcal{H}} z_h \sum_{k \in \mathcal{K}} \lambda_k \, \delta_h^{x_k \widetilde{\mu}_k} = \frac{1}{K} \sum_{k \in \mathcal{K}} \sum_{h \in \mathcal{H}} \lambda_k \, z_h \, \delta_h^{x_k \widetilde{\mu}_k} \\
&= \frac{1}{K} \sum_{k \in \mathcal{K}} \lambda_k \sum_{\substack{h \in \mathcal{H} \\ h \leq x_k \widetilde{\mu}_k}} z_h \underset{(*)}{=} \frac{1}{K} \sum_{\substack{k \in \mathcal{K} \\ x_k \widetilde{\mu}_k \geq 0}} \lambda_k \phi(x_k \, \widetilde{\mu}_k) = \frac{1}{K} \sum_{k \in \mathcal{M}(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu)} \lambda_k \, \phi |\widetilde{\mu}_k|.
\end{aligned}$$

Here, $(\diamond)$ is by (C6), while $(*)$ is by (C31), because $h \geq 0$ for all $h \in \mathcal{H}$. $\diamond$ `Claim 1`

Claim 2: *$F(\boldsymbol{\mathcal{C}}; \mu) \subseteq F_\phi(\boldsymbol{\mathcal{C}}; \mu)$ for all $\boldsymbol{\mathcal{C}} \in \langle \mathfrak{G}, \epsilon \rangle$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$.*

Proof: Let $\mathbf{x} \in F(\boldsymbol{\mathcal{C}}; \mu)$. For any other $\mathbf{y} \in \mathcal{X}$, defining formulae (C7) and (C8) yield $\mathbf{g}(\mathbf{y}, \boldsymbol{\mathcal{C}}, \mu) - \mathbf{g}(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu) \in \mathcal{D}_{\mathcal{H}}(F, \langle \mathfrak{G}, \epsilon \rangle)$, so $\mathbf{z} \bullet (\mathbf{g}(\mathbf{y}, \boldsymbol{\mathcal{C}}, \mu) - \mathbf{g}(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu)) \leq 0$, and hence $\mathbf{z} \bullet \mathbf{g}(\mathbf{y}, \boldsymbol{\mathcal{C}}, \mu) \leq \mathbf{z} \bullet \mathbf{g}(\mathbf{x}, \boldsymbol{\mathcal{C}}, \mu)$. Thus, $F(\boldsymbol{\mathcal{C}}; \mu) \subseteq \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} (\mathbf{z} \bullet \mathbf{g}(\mu, \boldsymbol{\mathcal{C}}, \mathbf{x}))$. Thus, Claim 1 and Proposition 4.3(a) imply that $F(\boldsymbol{\mathcal{C}}; \mu) \subseteq F_\phi(\boldsymbol{\mathcal{C}}; \mu)$, as desired. $\diamond$ `Claim 2`

Now, $\mathfrak{G} \subseteq \langle \mathfrak{G}, \epsilon \rangle$. Thus, Claim 2 implies that $F(\boldsymbol{\mathcal{G}}; \mu) \subseteq F_\phi(\boldsymbol{\mathcal{G}}; \mu)$ for all $\boldsymbol{\mathcal{G}} \in \mathfrak{G}$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{G}})$. Thus, Lemma C.6(b) says that $F(\boldsymbol{\mathcal{C}}; \mu) \subseteq F_\phi(\boldsymbol{\mathcal{C}}; \mu)$ for all $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$ and $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$, as desired.

(b) Let $\operatorname{cov}(F)$ be the set of additive majority rules which cover F on $\Delta_{\mathcal{H}}(\mathfrak{E})$. Part (a) implies that $\operatorname{cov}(F) \neq \emptyset$.

Claim 3: *$\operatorname{cov}(F)$ is finite.*

Proof: Recall that $\mathfrak{G}$ is finite. Suppose $\mathfrak{G} = \{\boldsymbol{\mathcal{C}}_1, \ldots, \boldsymbol{\mathcal{C}}_N\}$, where, for all $n \in [1 \ldots N]$, $\boldsymbol{\mathcal{C}}_n = (\mathcal{K}_n, \boldsymbol{\lambda}_n, \mathcal{X}_n, \mathcal{Y}_n)$. Now, $\mathcal{K}_n$ is finite and $\mathcal{Y}_n \subseteq \{\pm 1\}^{\mathcal{K}_n}$; thus, $\mathcal{Y}_n$ is also finite. Finally, $\mathcal{H}$ is finite, so $\Delta_{\mathcal{H}}(\mathcal{Y}_n)$ is finite. Thus, if $\mathcal{F}_n$ is the set of all possible judgement aggregation rules from $\Delta_{\mathcal{H}}(\mathcal{Y}_n)$ into $\mathcal{X}_n$, then $\mathcal{F}_n$ is also finite.

By Lemma C.6(b), any judgement aggregation rule F on $\Delta_{\mathcal{H}}(\mathfrak{E})$ which satisfies Combination and Scale invariance is obtained by choosing one rule $F_n \in \mathcal{F}_n$ for each $n \in [1 \ldots N]$. Thus, the set of all judgement aggregation rules on $\Delta_{\mathcal{H}}(\mathfrak{E})$ satisfying Combination and Scale invariance is also finite. By Lemma C.5, this means the set of all additive majority rules on $\Delta_{\mathcal{H}}(\mathfrak{E})$ is finite. Thus, $\operatorname{cov}(F)$ is finite. $\diamond$ `Claim 3`

Lemma C.7(b) implies that we can assume without loss of generality that all the rules in $\operatorname{cov}(F)$ have real-valued gain functions. Claim 3 implies that we can write $\operatorname{cov}(F) = \{F_{\phi_1}, F_{\phi_2}, \ldots, F_{\phi_L}\}$, for some gain functions $\phi_1, \phi_2, \ldots, \phi_L : [-1, 1] \longrightarrow \mathbb{R}$. Now define $\phi := \sum_{\ell=1}^{L} \phi_\ell$.

Claim 4: *F_ϕ covers F on $\Delta_{\mathcal{H}}(\mathfrak{E})$.*

Proof: Let $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y}) \in \mathfrak{E}$ and let $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$. Then for all $\ell \in [1 \ldots L]$, we have $F(\mathcal{C};\mu) \subseteq F_{\phi_\ell}(\mathcal{C};\mu)$, by definition of the set $\mathrm{cov}(F)$. This means: for all $\mathbf{x} \in F(\mathcal{C};\mu)$ and all other $\mathbf{y} \in \mathcal{X}$, we have $\phi_\ell(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{x} \geq \phi_\ell(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{y}$. By summing these inequalities over all $\ell \in [1 \ldots L]$, it follows that $\phi(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{x} \geq \phi(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{y}$. This holds for all $\mathbf{x} \in F(\mathcal{C};\mu)$ and all other $\mathbf{y} \in \mathcal{X}$; it follows that $F(\mathcal{C};\mu) \subseteq F_\phi(\mathcal{C};\mu)$, as desired. ◇ `Claim 4`

It remains to prove that F_ϕ is the *minimal* rule in $\mathrm{cov}(F)$.

Claim 5: *For all $\ell \in [1 \ldots L]$, the rule F_{ϕ_ℓ} covers F_ϕ on $\Delta_{\mathcal{H}}(\mathfrak{E})$.*

Proof: (by contrapositive) Let $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y}) \in \mathfrak{E}$, let $\mu \in \Delta_{\mathcal{H}}(\mathcal{Y})$, and suppose $\mathbf{x} \notin F_{\phi_\ell}(\mathcal{C};\mu)$ for some $\ell \in [1 \ldots L]$. We will show that $\mathbf{x} \notin F_\phi(\mathcal{C};\mu)$.

By hypothesis F is total, so there exists some $\mathbf{y} \in F(\mathcal{C};\mu)$. Then for all $\ell' \in [1 \ldots L]$, we have $\mathbf{y} \in F_{\phi_{\ell'}}(\mathcal{C};\mu)$, and thus, $\phi_{\ell'}(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{y} \geq \phi_{\ell'}(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{x}$. In particular, $\mathbf{y} \in F_{\phi_\ell}(\mathcal{C};\mu)$, whereas $\mathbf{x} \notin F_{\phi_\ell}(\mathcal{C};\mu)$; thus, $\phi_\ell(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{y} > \phi_\ell(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{x}$. By summing these inequalities over all $\ell' \in [1 \ldots L]$, it follows that $\phi(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{y} > \phi(\widetilde{\boldsymbol{\mu}}) \underset{\lambda}{\bullet} \mathbf{x}$. Thus, $\mathbf{x} \notin F_\phi(\mathcal{C};\mu)$. ◇ `Claim 5`

Claims 4 and 5 imply that F_ϕ is a minimal covering of F by an additive majority rule. If F is total, then clearly minimality implies uniqueness.

(c) This follows immediately from Lemma C.7. □

Proof of Theorem C.1. The proof is identical to the proof of Theorem C.2, except that the first paragraph invokes Lemma C.3(a,b) instead of Lemma C.3(a,c). Likewise, later steps in the proof invoke Lemmas C.6(a) and C.7(a) instead of Lemmas C.6(b) and C.7(b). Notice that we do not need F to satisfy Scale invariance to invoke any of these lemmas, or at any other stage in the proof. □

D Proofs of the main results in Section 5

Theorem 1 is obtained by using an ultrapower construction to "stitch together" the gain functions defined in Theorems C.1 and C.2 for every possible choice of weight function $\mathcal{H}$ and every finite sub-environment of $\mathfrak{E}$.

Proof of Theorem 1. Lemma C.5 says that any additive majority rule satisfies Supermajority efficiency, Combination and Scale invariance. This proves part (a). It remains to prove parts (b)-(d). Throughout the rest of the proof, we must distinguish between two cases:

Case 1. $\mathfrak{E}$ is a rationally weighted judgement environment.

Case 2. $\mathfrak{E}$ is a minimally rich judgement environment, and F satisfies Scale invariance.

We will use Theorem C.1 to handle Case 1, and and Theorem C.2 to handle Case 2.

(b) "$\Longleftarrow$" exactly the same proof as Theorem C.2.

"$\Longrightarrow$" Let $\wp(\mathfrak{E})$ be the set of all finite subsets of $\mathfrak{E}$ (in Case 1) or finitely generated sub-environments of $\mathfrak{E}$ (in Case 2); we will denote a generic element of $\wp(\mathfrak{E})$ by $\mathfrak{D}$. Note that in Case 2, we have $\langle\mathfrak{D}\rangle = \mathfrak{D}$ for any $\mathfrak{D} \in \wp(\mathfrak{E})$, by condition (FG1). Let $\mathcal{I} := \wp[0,1] \times \wp(\mathfrak{E})$. Thus, $\mathcal{I}$ is a collection of possible "inputs" to Theorem C.1 (in Case 1) or Theorem C.2 (in Case 2).

For any finite collection $\mathcal{T} := \{(\boldsymbol{\mathcal{C}}_1, \mu_1), \ldots, (\boldsymbol{\mathcal{C}}_N, \mu_N)\} \subset \Delta\langle\mathfrak{E}\rangle$, define $\mathcal{I}_\mathcal{T} := \{(\mathcal{H}, \mathfrak{D}) \in \mathcal{I};\ \boldsymbol{\mathcal{C}}_n \in \langle\mathfrak{D}\rangle$ and $\mu_n \in \Delta_\mathcal{H}(\boldsymbol{\mathcal{C}}_n)$ for all $n \in [1 \ldots N]\}$. Then let $\mathfrak{F} := \{\mathcal{J} \subseteq \mathcal{I};\ \mathcal{I}_\mathcal{T} \subseteq \mathcal{J}$ for some nonempty finite $\mathcal{T} \subset \Delta\langle\mathfrak{E}\rangle\}$.

Claim 1: *$\mathfrak{F}$ is a free filter.*

Proof: We must check axioms (F0)-(F2) from Appendix A.

(F0) Let $\mathcal{T} := \{(\boldsymbol{\mathcal{C}}_1, \mu_1), \ldots, (\boldsymbol{\mathcal{C}}_N, \mu_N)\} \subset \Delta\langle\mathfrak{E}\rangle$. Let $\mathcal{H} \in \wp[0,1]$ be arbitrary, and suppose $\mathrm{supp}(\mathcal{H}) = [1 \ldots W]$ for some $W \in \mathbb{N}$. For all $n \in [1 \ldots N]$, let $\boldsymbol{\mathcal{C}}_n := (\mathcal{K}_n, \boldsymbol{\lambda}_n, \mathcal{X}_n, \mathcal{Y}_n)$. Let $M := W \cdot |\mathcal{Y}_1| \cdot |\mathcal{Y}_2| \cdots |\mathcal{Y}_N|$, and let $\beta : [1 \ldots M] \longrightarrow [1 \ldots W] \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_N$ be some bijection. Define $\mathcal{H}' \in \wp[0,1]$ as follows: for all $m \in [1 \ldots M]$, if $\beta(m) = (w, \mathbf{y}_1, \ldots, \mathbf{y}_N)$, then let $\mathcal{H}'(m) := \mathcal{H}(w) \cdot \mu_1(\mathbf{y}_1) \cdots \mu_N(\mathbf{y}_N)$. It is easy to check that $(\boldsymbol{\mathcal{C}}_n, \mu_n) \in \Delta_{\mathcal{H}'}(\boldsymbol{\mathcal{C}}_n)$ for all $n \in [1 \ldots N]$. Let $\mathfrak{D} \subseteq \mathfrak{E}$ be any finite subset such that $\boldsymbol{\mathcal{C}}_1, \ldots, \boldsymbol{\mathcal{C}}_N \in \langle\mathfrak{D}\rangle$. (For example, let $\mathfrak{D} = \{\boldsymbol{\mathcal{C}}_1, \ldots, \boldsymbol{\mathcal{C}}_N\}$.) Then $(\boldsymbol{\mathcal{C}}_n, \mu_n) \in \Delta_{\mathcal{H}'}\langle\mathfrak{D}\rangle$ for all $n \in [1 \ldots N]$. Thus, $(\mathcal{H}', \mathfrak{D}) \in \mathcal{I}_\mathcal{T}$.

We can repeat this construction for any $\mathcal{H} \in \wp[0,1]$; thus, $\mathcal{I}_\mathcal{T}$ is infinite. This holds for any finite $\mathcal{T} \subset \Delta\langle\mathfrak{E}\rangle$. Every element of $\mathfrak{F}$ must contain $\mathcal{I}_\mathcal{T}$ for some finite $\mathcal{T} \subset \Delta\langle\mathfrak{E}\rangle$; thus, every element of $\mathfrak{F}$ is infinite.

(F1) Let $\mathcal{E}, \mathcal{F} \in \mathfrak{F}$. Then there exist finite sets $\mathcal{S}, \mathcal{T} \subset \Delta\langle\mathfrak{E}\rangle$ such that $\mathcal{I}_\mathcal{S} \subseteq \mathcal{E}$ and $\mathcal{I}_\mathcal{T} \subseteq \mathcal{F}$. But then $\mathcal{S} \cup \mathcal{T}$ is also finite, and $\mathcal{I}_{\mathcal{S}\cup\mathcal{T}} = \mathcal{I}_\mathcal{S} \cap \mathcal{I}_\mathcal{T} \subseteq \mathcal{E} \cap \mathcal{F}$; thus, $\mathcal{E} \cap \mathcal{F} \in \mathfrak{F}$.

(F2) Suppose $\mathcal{E} \in \mathfrak{F}$ and $\mathcal{E} \subseteq \mathcal{D}$. Then there is some finite $\mathcal{T} \subset \Delta\langle\mathfrak{E}\rangle$ such that $\mathcal{I}_\mathcal{T} \subseteq \mathcal{E}$. But then $\mathcal{I}_\mathcal{T} \subseteq \mathcal{D}$; thus $\mathcal{D} \in \mathfrak{F}$ also. ◇ `Claim 1`

Now Claim 1 and the Ultrafilter Lemma yields a free ultrafilter $\mathfrak{U}$ with $\mathfrak{F} \subseteq \mathfrak{U}$. Let ${}^*\mathbb{R}$ be the hyperreal field defined by $\mathfrak{U}$; in other words, ${}^*\mathbb{R} := \mathbb{R}^\mathcal{I}/\underset{\mathfrak{U}}{\approx}$, where $\underset{\mathfrak{U}}{\approx}$ is the equivalence relation on $\mathbb{R}^\mathcal{I}$ defined by $\mathfrak{U}$. (See Appendix A for details.) We define $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ as follows. For all $(\mathcal{H}, \mathfrak{D}) \in \mathcal{I}$, we can apply Theorem C.1(a) (in Case 1) or Theorem C.2(a) (in Case 2) to obtain an odd, strictly increasing function $\phi_{\mathcal{H},\mathfrak{D}} : \mathcal{H} \longrightarrow \mathbb{R}$ such that the associated additive majority rule $F_{\phi_{\mathcal{H},\mathfrak{D}}}$ covers F on $\Delta_\mathcal{H}\langle\mathfrak{D}\rangle$. Recall that $\mathcal{H}$ is a finite subset of $[-1,1]$ (because $\mathcal{H}$ has finite support). Thus, we can extend $\phi_{\mathcal{H},\mathfrak{D}}$ to an odd, continuous increasing function $\phi_{\mathcal{H},\mathfrak{D}} : [-1,1] \longrightarrow \mathbb{R}$, by linearly interpolating the values between the points in $\mathcal{H}$. Now, for any $r \in [-1,1]$, define $\widehat{\phi}(r) \in \mathbb{R}^\mathcal{I}$ by:

$$\widehat{\phi}(r)_{(\mathcal{H},\mathfrak{D})} := \phi_{\mathcal{H},\mathfrak{D}}(r), \quad \text{for all } (\mathcal{H}, \mathfrak{D}) \in \mathcal{I}. \tag{E1}$$

Then define $\phi(r) \in {}^*\mathbb{R}$ to be the $\underset{\mathfrak{U}}{\approx}$-equivalence class of $\widehat{\phi}(r)$.

Claim 2: *ϕ is odd and strictly increasing.*

Proof: Odd. Let $r \in [-1,1]$. For all $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}$, we have $\phi_{\mathcal{H},\mathfrak{D}}(-r) = -\phi_{\mathcal{H},\mathfrak{D}}(r)$, because $\phi_{\mathcal{H},\mathfrak{D}}$ is odd by construction. But $\mathcal{I} \in \mathfrak{U}$, by definition of $\mathfrak{U}$. Thus, $\phi(-r) = -\phi(r)$.

Increasing. Let $h, i \in [-1,1]$, with $h < i$. For all $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}$, we have $\phi_{\mathcal{H},\mathfrak{D}}(h) < \phi_{\mathcal{H},\mathfrak{D}}(i)$, because $\phi_{\mathcal{H},\mathfrak{D}}$ is increasing by construction. But $\mathcal{I} \in \mathfrak{U}$ by definition of $\mathfrak{U}$. Thus, we obtain $\phi(h) < \phi(i)$, by defining formula (A1). $\diamond$ `Claim 2`

Claim 3: *For any $\mathcal{C} \in \langle\mathfrak{E}\rangle$ and $\mu \in \Delta(\mathcal{C})$, we have $F(\mathcal{C};\mu) \subseteq F_\phi(\mathcal{C};\mu)$.*

Proof: Suppose $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. Let $\mu \in \Delta(\mathcal{Y})$ and let $\mathbf{x} \in F(\mathcal{C};\mu)$; we must show that $\mathbf{x} \in F_\phi(\mathcal{C};\mu)$. For all $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}_{\{(\mathcal{C};\mu)\}}$, Theorem C.1(a) (in Case 1) or Theorem C.2(a) (in Case 2) says that $F(\mathcal{C};\mu) \subseteq F_{\phi_{\mathcal{H},\mathfrak{D}}}(\mathcal{C};\mu)$; thus, for all $\mathbf{y} \in \mathcal{X}$, we have $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) \geq 0$. Thus, if we define $\mathcal{I}_\mathbf{y} := \{(\mathcal{H},\mathfrak{D}) \in \mathcal{I};\ (\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) \geq 0\}$, then $\mathcal{I}_\mathbf{y} \supseteq \mathcal{I}_{\{(\mathcal{C};\mu)\}}$. But $\mathcal{I}_{\{(\mathcal{C};\mu)\}} \in \mathfrak{U}$; thus, $\mathcal{I}_\mathbf{y} \in \mathfrak{U}$, by axiom (F2) from Appendix A. (In other words: $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) \geq 0$ for "almost all" $(\mathcal{H},\mathfrak{D})$ in $\mathcal{I}$.) Thus, $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) \geq 0$, by defining formula (A1). This holds for all $\mathbf{y} \in \mathcal{X}$; thus, $\mathbf{x} \in F_\phi(\mathcal{C};\mu)$. This holds for all $\mathbf{x} \in F(\mathcal{C};\mu)$; thus, $F(\mathcal{C};\mu) \subseteq F_\phi(\mathcal{C};\mu)$. $\diamond$ `Claim 3`

Now let $G := F_\phi$ to prove part (b).

(c) If F is total, then $F(\mathcal{C};\mu)$ is always nonempty. Thus, $F(\mathcal{C};\mu) = F_\phi(\mathcal{C};\mu)$ whenever $F_\phi(\mathcal{C};\mu)$ is single-valued. But ϕ is strictly increasing, so F_ϕ is single-valued on a dense, open subset of $\Delta(\mathcal{Y})$, by Proposition 4.4(b).

(d) Suppose F is total. For all $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}$, Theorem C.1(b) (in Case 1) or Theorem C.2(b) (in Case 2) says that we can choose the gain functions $\phi_{\mathcal{H},\mathfrak{D}}$ such that the additive majority rule $F_{\phi_{\mathcal{H},\mathfrak{D}}}$ is the unique *minimal* covering of F on $\Delta_\mathcal{H}(\mathfrak{D})$. Suppose we perform the construction in part (b) using this set of gain functions; we claim the resulting rule F_ϕ is a minimal covering of F on $\Delta\langle\mathfrak{E}\rangle$.

To see this, let $\psi : [-1,1] \longrightarrow {}^*\mathbb{R}$ be another gain function, such that the additive majority rule F_ψ covers F on $\Delta\langle\mathfrak{E}\rangle$. Let $\mathcal{C} \in \langle\mathfrak{E}\rangle$ and let $\mu \in \Delta(\mathcal{C})$; we must show that $F_\phi(\mathcal{C};\mu) \subseteq F_\psi(\mathcal{C};\mu)$. So, let $\mathbf{x} \in F_\phi(\mathcal{C};\mu)$; we will show that $\mathbf{x} \in F_\psi(\mathcal{C};\mu)$.

Suppose $\mathcal{C} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. For all $\mathbf{y} \in \mathcal{X}$, let $\mathcal{I}_{\mathbf{x},\mathbf{y}} := \{(\mathcal{H},\mathfrak{D}) \in \mathcal{I};\ \mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) \geq \mathbf{y} \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}})\}$. Then $\mathcal{I}_{\mathbf{x},\mathbf{y}} \in \mathfrak{U}$, by formula (A1) (because $\mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) \geq \mathbf{y} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}})$, because $\mathbf{x} \in F_\phi(\mathcal{C};\mu)$). Thus, if we define

$$\mathcal{I}_\mathbf{x} \;:=\; \left\{(\mathcal{H},\mathfrak{D}) \in \mathcal{I}_{\{(\mathcal{C};\mu)\}}\,;\ \mathbf{x} \in F_{\phi_{\mathcal{H},\mathfrak{D}}}(\mathcal{C};\mu)\right\} \quad = \quad \mathcal{I}_{\{(\mathcal{C};\mu)\}} \cap \bigcap_{\mathbf{y}\in\mathcal{X}} \mathcal{I}_{\mathbf{x},\mathbf{y}},$$

then $\mathcal{I}_\mathbf{x} \in \mathfrak{U}$, by axiom (F1) (because it is a finite intersection of $\mathfrak{U}$-elements, because $\mathcal{X}$ is a finite set). Thus, $\mathcal{I}_\mathbf{x}$ is nonempty, by axiom (F0).

Now, for all $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}_{\{(\mathcal{C};\mu)\}}$, the rule F_ψ covers F on $\Delta_\mathcal{H}(\mathfrak{D})$. Thus, F_ψ also covers $F_{\phi_{\mathcal{H},\mathfrak{D}}}$ (because $F_{\phi_{\mathcal{H},\mathfrak{D}}}$ is the *minimal* covering of F on $\Delta_\mathcal{H}(\mathfrak{D})$, by hypothesis). Thus, if we take any $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}_\mathbf{x}$, we obtain $\mathbf{x} \in F_{\phi_{\mathcal{H},\mathfrak{D}}}(\mathcal{C};\mu) \subseteq F_\psi(\mathcal{C};\mu)$. Thus, $\mathbf{x} \in F_\psi(\mathcal{C};\mu)$.

This argument works for all $\mathbf{x} \in F_\phi(\mathcal{C};\mu)$. Thus, $F_\phi(\mathcal{C};\mu) \subseteq F_\psi(\mathcal{C};\mu)$, as desired. $\Box$

The proof of Theorem 2 uses the following result.

Lemma E.1 *Let $\mathcal{X} \subseteq \mathcal{Y} \subseteq \{\pm 1\}^{\mathcal{K}}$, and let $F, G : \Delta(\mathcal{Y}) \rightrightarrows \mathcal{X}$ be two judgement aggregation rules. Suppose $F(\mu) \subseteq G(\mu)$ for all $\mu \in \Delta(\mathcal{Y})$, and G is monotone, and F is continuous and total. Then $F(\mu) = G(\mu)$ for all $\mu \in \Delta(\mathcal{Y})$.*

Proof: Let $\mu \in \Delta(\mathcal{Y})$. We have $F(\mu) \subseteq G(\mu)$ by hypothesis; we must show $F(\mu) \supseteq G(\mu)$. So, let $\mathbf{x} \in G(\mu)$; we will show that $\mathbf{x} \in F(\mu)$. Let $\delta_{\mathbf{x}} \in \Delta(\mathcal{Y})$ be the unanimous profile at $\mathbf{x}$. For all $n \in \mathbb{N}$, define $\mu_n := (1 - \frac{1}{n})\mu + \frac{1}{n}\delta_{\mathbf{x}}$. Then μ_n is more supportive of $\mathbf{x}$ than μ, so $G(\mu_n) = \{\mathbf{x}\}$ because G is monotone. Thus, $F(\mu_n) = \{\mathbf{x}\}$ because $F \subseteq G$, and $F(\mu_n)$ must be nonempty because F is total. However, $\lim_{n\to\infty} \mu_n = \mu$, and F is continuous. Thus, $\mathbf{x} \in F(\mu)$, as desired. $\square$

Proof of Theorem 2. "$\Longrightarrow$" This follows from Theorem 1(a).

"$\Longleftarrow$" Let ϕ be the gain function from Theorem 1(b). Then F_ϕ covers F. The rule F_ϕ is monotone, by Proposition 4.4(a). Meanwhile F is continuous and total by hypothesis. Thus, Lemma E.1 implies that $F(\mu) = F_\phi(\mu)$ for all $\mu \in \Delta(\mathcal{Y})$. $\square$

Proof of Proposition 5.3. Suppose $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. Let $\{\mu^n\}_{n=1}^\infty \subset \Delta(\mathcal{Y})$ and $\mu \in \Delta(\mathcal{Y})$, and suppose $\lim_{n\to\infty} \mu^n = \mu$. Let $\mathbf{x} \in \mathcal{X}$, and suppose $\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}}, \mu^n)$ for all $n \in \mathbb{N}$; we must show that $\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}}, \mu)$. Let $\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{x}\}$. Then

$$\begin{aligned}
&\left(\lim_{n\to\infty} \mu^n = \mu\right) \underset{(*)}{\Longrightarrow} \left(\lim_{n\to\infty} \widetilde{\mu}^n = \widetilde{\mu}\right) \underset{(\dagger)}{\Longrightarrow} \left(\lim_{n\to\infty} \phi(\widetilde{\mu}^n_k) = \phi(\widetilde{\mu}_k), \text{ for all } k \in \mathcal{K}\right) \\
&\Longrightarrow \left(\lim_{n\to\infty} \sum_{k\in\mathcal{K}} \lambda_k \, (x_k - y_k) \, \phi(\widetilde{\mu}^n_k) = \sum_{k\in\mathcal{K}} \lambda_k \, (x_k - y_k) \, \phi(\widetilde{\mu})\right) \\
&\Longleftrightarrow \left(\lim_{n\to\infty} (\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\mu}^n) \; = \; (\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\mu})\right) \underset{(\ddagger)}{\Longrightarrow} \left((\mathbf{x} - \mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\mu}) \geq 0\right).
\end{aligned}$$

Here, $(*)$ is because the map $\Delta(\mathcal{Y}) \ni \mu \mapsto \widetilde{\mu} \in \mathbb{R}^{\mathcal{K}}$ is continuous, and $(\dagger)$ is because ϕ is continuous. Next, $(\ddagger)$ is because $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\mu}^n) \geq 0$ for all $n \in \mathbb{N}$, because $\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}}, \mu^n)$. Thus, $\mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\mu}) \geq \mathbf{y} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\mu})$ for all $\mathbf{y} \in \mathcal{X}$. Thus, $\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}}, \mu)$ as desired. $\square$

Proof of Theorem 3. This follows from Theorem C.1; let $\mathcal{H} := \{\frac{m}{N};\; m \in [0 \ldots N]\}$. $\square$

F More on representation by ${}^*\mathbb{R}$-valued gain functions

First, we demonstrate two statements made at the end of Section 5 showing why a real-valued representation of the Leximin rule is not possible in general, and how a hyper-real valued representation overcomes this difficulty.

Proposition F.1 *Let $\mathfrak{E}$ be any judgement environment.*

(a) *There is no real-valued gain function ϕ such that* Leximin $= F_\phi$ *on* $\Delta_N(\mathfrak{E})$ *for all* $N \in \mathbb{N}$.

(b) *If ω is an infinite hyperreal number, then H^ω is well-defined and equal to* Leximin.

Proof: (a) (by contradiction) Suppose Leximin $= F_\phi$ for some real-valued gain function ϕ, and consider a population of size N. In the Leximin rule, a single majority of M voters overrules any number of majorities of $M-1$ voters. Thus, we must have $\phi(\frac{M}{N}) > 2 \cdot \phi(\frac{M-1}{N})$. If $q := \frac{M}{N}$ and $\epsilon = \frac{1}{N}$, then we get $\phi(q) > 2 \cdot \phi(q - \epsilon)$. But if N can be arbitrarily large, then ϵ can be arbitrarily small, while q can be any rational number in $[0, 1]$. Thus, we obtain $\phi(q) > 2 \cdot \phi(q')$ for all rationals $q' < q$ in $[0, 1]$. Clearly, it is impossible for any real-valued function to behave in this way.

(b) For any $r, s \in [0, 1]$, if $r > s$ and ω is infinite, then the ratio r^ω / s^ω is infinite (because it must be larger than $(r/s)^N$ for any $N \in \mathbb{N}$). Thus, for any judgement context $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$ in $\mathfrak{E}$, any $\mu \in \Delta(\mathcal{Y})$, and any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $\mathbf{x} \underset{\mu}{\succ} \mathbf{y}$, then $\phi(\widetilde{\boldsymbol{\mu}}) \underset{\boldsymbol{\lambda}}{\bullet} \mathbf{x} > \phi(\widetilde{\boldsymbol{\mu}}) \underset{\boldsymbol{\lambda}}{\bullet} \mathbf{y}$; hence $\mathbf{y} \notin F_\phi(\boldsymbol{\mathcal{C}}, \mu)$. This shows that $F_\phi(\boldsymbol{\mathcal{C}}, \mu) \subseteq \text{Leximin}(\boldsymbol{\mathcal{C}}, \mu)$. On the other hand, for any $\mathbf{x}, \mathbf{y} \in \text{Leximin}(\boldsymbol{\mathcal{C}}, \mu)$, we must have $\mathbf{x} \underset{\mu}{\approx} \mathbf{y}$, which means $\gamma_{\mu,\mathbf{x}} = \gamma_{\mu,\mathbf{y}}$, which means $\mathbf{x} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) = \mathbf{y} \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}})$. Thus $F_\phi(\boldsymbol{\mathcal{C}}, \mu) = \text{Leximin}(\boldsymbol{\mathcal{C}}, \mu)$. □

The alert reader will have noticed that we do not need the full structure of ${}^*\mathbb{R}$ to define additive majority rules, or to obtain some of our auxiliary results. Since, for these purposes, we only multiply by reals, not by hyperreals, we can make do with a much thinner structure. A *linearly ordered vector space* is a triple $(\mathbb{L}, +, \cdot, >)$, where $\mathbb{L}$ is a set, $+$ is binary operation making $\mathbb{L}$ into an abelian group (i.e. $+$ is associative, commutative, and invertible), $\cdot$ is a "scalar multiplication" operation of $\mathbb{R}$ on $\mathbb{L}$ which distributes over $+$, and $>$ is a linear ordering relation compatible with $+$, and $\cdot$: for all $\ell, m \in \mathbb{L}$ and $r \in \mathbb{R}$, we have $\ell > 0$ iff $\ell + m > m$, and $r\,\ell > 0$ if $r > 0$ and $\ell > 0$, or if $r < 0$ and $\ell < 0$. For example, $\mathbb{R}$ and ${}^*\mathbb{R}$ are both linearly ordered vector spaces. Another example is $\mathbb{R}^N$ with the lexicographical ordering. An $\mathbb{L}$*-valued gain function* is now any increasing function $\phi : [-1, 1] \longrightarrow \mathbb{L}$. Given any $\mathbb{L}$-valued gain function ϕ, we can define the *additive majority rule* F_ϕ as in equation (4.7).

At first it might appear that this definition provides greater generality than the hyperreal-valued gain functions considered in the text. However, this extra generality is illusory. Indeed, the next result says that one can make do with the hyperreal field that is already given by the ultrapower construction in the proof of Theorem 1.

Proposition F.2 *Let $\mathbb{L}$ be any linearly ordered vector space, and let $\psi : [-1,1] \longrightarrow \mathbb{L}$ be any gain function. Then there exists a hyperreal field ${}^*\mathbb{R}$ and a gain function $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ such that, for any judgement problem $(\boldsymbol{\mathcal{C}};\mu)$* (weighted or unweighted)*, $F_\phi(\boldsymbol{\mathcal{C}};\mu) = F_\psi(\boldsymbol{\mathcal{C}};\mu)$.*

The proof requires the following variant of Lemma C.7.

Lemma F.3 *Let $\mathbb{L}$ be a linearly ordered vector space, and let $\psi : [-1,1] \longrightarrow \mathbb{L}$ be a gain function. Let $\mathfrak{E} = \langle \mathfrak{G} \rangle$ for some finite collection $\mathfrak{G}$ of judgement contexts. For any $\mathcal{H} \in \wp[0,1]$, there exists a real-valued gain function $\phi_{\mathcal{H}} : \mathcal{H} \longrightarrow \mathbb{R}$ such that $F_{\phi_{\mathcal{H}}}(\boldsymbol{\mathcal{C}};\mu) = F_\psi(\boldsymbol{\mathcal{C}};\mu)$ for all $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$ and all $\mu \in \Delta_{\mathcal{H}}(\boldsymbol{\mathcal{C}})$.*

Proof: The proof is identical to the proof of Lemma C.7(a) —simply replace ${}^*\mathbb{R}$ with $\mathbb{L}$ everywhere. $\square$

Proof of Proposition F.2. The proof strategy is almost identical to the proof of Theorem 1(b). Let $\mathfrak{E}$ be the set of *all* judgement contexts. (Thus, $\mathfrak{E} = \langle \mathfrak{E} \rangle$.) Define $\mathcal{I}$ as in the proof of Theorem 1, and for each $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}$, let $\phi_{\mathcal{H},\mathfrak{D}} : \mathcal{H} \longrightarrow \mathbb{R}$ be the gain function from Lemma F.3. Then define $\phi : [-1,1] \longrightarrow {}^*\mathbb{R}$ as in Eq.(E1). The proof of Claim 2 is exactly as before. But we replace Claim 3 with the following:

Claim 3′: *For any $\boldsymbol{\mathcal{C}} \in \mathfrak{E}$ and $\mu \in \Delta(\boldsymbol{\mathcal{C}})$, we have $F_\phi(\boldsymbol{\mathcal{C}};\mu) = F_\psi(\boldsymbol{\mathcal{C}};\mu)$.*

To prove Claim 3′, let $\boldsymbol{\mathcal{C}} = (\mathcal{K}, \boldsymbol{\lambda}, \mathcal{X}, \mathcal{Y})$. Suppose $\mathbf{x} \in F_\psi(\boldsymbol{\mathcal{C}};\mu)$ and $\mathbf{z} \in \mathcal{X} \setminus F_\psi(\boldsymbol{\mathcal{C}};\mu)$. We must show that $\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}};\mu)$ and $\mathbf{z} \notin F_\phi(\boldsymbol{\mathcal{C}};\mu)$. For all $(\mathcal{H},\mathfrak{D}) \in \mathcal{I}_{\{(\boldsymbol{\mathcal{C}};\mu)\}}$, Lemma F.3 says that $F_\psi(\boldsymbol{\mathcal{C}};\mu) = F_{\phi_{\mathcal{H},\mathfrak{D}}}(\boldsymbol{\mathcal{C}};\mu)$; thus, for all $\mathbf{y} \in \mathcal{X}$, we have $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) \geq 0$, and furthermore, $(\mathbf{x}-\mathbf{z}) \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) > 0$. Thus, if we define $\mathcal{I}_{\mathbf{y},\geq} := \{(\mathcal{H},\mathfrak{D}) \in \mathcal{I};\ (\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) \geq 0\}$, then $\mathcal{I}_{\mathbf{y},\geq} \supseteq \mathcal{I}_{\{(\boldsymbol{\mathcal{C}};\mu)\}}$. Also, if we define $\mathcal{I}_{\mathbf{z},>} := \{(\mathcal{H},\mathfrak{D}) \in \mathcal{I};\ (\mathbf{x}-\mathbf{z}) \underset{\boldsymbol{\lambda}}{\bullet} \phi_{\mathcal{H},\mathfrak{D}}(\widetilde{\boldsymbol{\mu}}) > 0\}$, then $\mathcal{I}_{\mathbf{z},>} \supseteq \mathcal{I}_{\{(\boldsymbol{\mathcal{C}};\mu)\}}$. But $\mathcal{I}_{\{(\boldsymbol{\mathcal{C}};\mu)\}} \in \mathfrak{U}$; thus, we get $\mathcal{I}_{\mathbf{y},\geq} \in \mathfrak{U}$ and $\mathcal{I}_{\mathbf{z},>} \in \mathfrak{U}$, by axiom (F2) from Appendix A. Thus, $(\mathbf{x}-\mathbf{y}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) \geq 0$, by the defining formula (A1). This holds for all $\mathbf{y} \in \mathcal{X}$; thus, $\mathbf{x} \in F_\phi(\boldsymbol{\mathcal{C}};\mu)$. Likewise, $(\mathbf{x}-\mathbf{z}) \underset{\boldsymbol{\lambda}}{\bullet} \phi(\widetilde{\boldsymbol{\mu}}) > 0$ by the defining formula (A1), so $\mathbf{z} \notin F_\phi(\boldsymbol{\mathcal{C}};\mu)$. $\square$

References

Balinski, M., Laraki, R., 2010. Majority Judgment: Measuring, Ranking, and Electing. MIT Press, Boston, MA.

Bandelt, H., Barthelemy, J., 1984. Medians in median graphs. Discrete Applied Mathematics, 131–142.

Barberà, S., Masso, J., Neme, A., 1997. Voting under constraints. Journal of Economic Theory 76, 298–321.

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

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

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

Bossert, W., Sprumont, Y., 2014. Strategyproof preference aggregation. Games and Economic Behavior 85, 109–126.

Cedroni, L., Garzia, D. (Eds.), 2010. Voting Advice Applications - The State of the Art. Scriptaweb, Napoli.

Christiano, T., 2008. Democracy. In: Zalta, E. N. (Ed.), The Stanford Encyclopedia of Philosophy, fall 2008 Edition. Metaphysics Research Lab, Stanford University.
URL `http://plato.stanford.edu/archives/fall2008/entries/democracy/`

Clifford, A. H., 1954. Note on Hahn's theorem on ordered abelian groups. Proc. Amer. Math. Soc. 5, 860–863.

Condorcet, M. d., 1785. An essay on the application of analysis to the probability of decisions rendered by the plurality of votes. In: McLean and Urken (1995), pp. 91–112.

Condorcet, M. d., 1788. On the constitution and the functions of provincial assemblies. In: McLean and Urken (1995), pp. 113–144.

Cusanus, N., 1434. On Catholic harmony, Book III. In: McLean and Urken (1995), Ch. 37, pp. 77–80.

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

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

Estlund, D., 2008. Democratic Authority. Princeton UP, Princeton and Oxford.

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

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

Goldblatt, R., 1998. Lectures on the hyperreals. Vol. 188 of Graduate Texts in Mathematics. Springer-Verlag, New York, an introduction to nonstandard analysis.

Goodin, R., Spiekermann, K., 2018. An Epistemic Theory of Democracy. Oxford University Press, Oxford, UK.

Gravett, K. A. H., 1956. Ordered abelian groups. Quart. J. Math. Oxford Ser. (2) 7, 57–63.

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.

Halpern, J. Y., 2010. Lexicographic probability, conditional probability, and nonstandard probability. Games Econom. Behav. 68 (1), 155–179.

Hausner, M., Wendel, J. G., 1952. Ordered vector spaces. Proc. Amer. Math. Soc. 3, 977–982.

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.

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

List, C., Pettit, P., 2011. Group Agency. Oxford University Press, Oxford, UK.

Lull, R., 1299. The art of elections. In: McLean and Urken (1995), pp. 73–76.

Manin, B., 1995. The Principles of Representative Government. Cambridge University Press, Cambridge, UK.

McKelvey, R., 1986. Covering, dominance, and institution-free properties of social choice. Amer. J. Polit. Sci. 30, 283–314.

McLean, I., Urken, A. (Eds.), 1995. Classics of Social Choice. Michigan University Press.

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.

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. The median rule in judgement aggregation. (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.

Nehring, K., Puppe, C., 2010. Abstract Arrowian aggregation. J. Econom. Theory 145 (2), 467–494.

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

Pivato, M., 2013a. Additive representation of separable preferences over infinite products. to appear in Theory and Decision.

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

Pivato, M., 2017. Epistemic democracy with correlated voters. J. Math. Econom. 72, 51–69.

Schwartzberg, M., 2015. Epistemic democracy and its challenges. Annual Review of Political Science 18, 187–203.

Slater, P., 1961. Inconsistencies in a schedule of paired comparisons. Biometrica 48, 303–312.

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

Tangian, A., 2014. Mathematical Theory of Democracy. Springer, Heidelberg.

Tideman, T. N., 1987. Independence of clones as a criterion for voting rules. Soc. Choice Welf. 4 (3), 185–206.

Waldron, J., 1999. The Dignity of Legislation. Cambridge University Press.

Young, H. P., 1975. Social choice scoring functions. SIAM J. Appl. Math. 28, 824–838.

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

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

Zavist, T. M., Tideman, T. N., 1989. Complete independence of clones in the ranked pairs rule. Soc. Choice Welf. 6 (2), 167–173.