Modeling individual expertise in group judgments

July 12, 2013

Abstract

Group judgments are often—implicitly or explicitly—influenced by their members' individual expertise. However, given that expertise is seldom recognized fully and that some distortions may occur (bias, correlation, etc.), it is not clear that differential weighting is an epistemically advantageous strategy with respect to straight averaging. Our paper characterizes a wide set of conditions under which differential weighting outperforms straight averaging and embeds the results into the multidisciplinary group decision-making literature.

1 Introduction

Groups frequently make judgments that are based on aggregating the opinions of its individual members. A panel of market analysts at Apple or Samsung may estimate the expected number of sales of a newly developed cell phone. A group of conservation biologists may assess the population size of a particular species in a specific habitat. A research group at the European Central may evaluate the merits of a particular monetary policy. Generally, such problems occur in any context where groups have to combine various opinions into a single group judgment (for a review paper, see Clemen 1989).

Even in cases of fully shared information, the assessment of the evidence will generally vary among the agents and depend on factors such as professional training, familiarity with similar situations in the past, and personal attitude toward the results. Thus, it will not come as a surprise that the individual judgments may differ. But how shall they be aggregated?

Often, some group members are more competent than others. Recognizing these *experts* may then become a crucial issue for improving group performance. Research in social psychology and management science has investigated the ability of humans to properly assess the expertise of other group members in such contexts (Clemen 1989; Bonner, Baumann and Dalal

2002; Larrick, Burson and Soll 2007). Most of this research stresses that recognizing experts is no easy task: perceived and actual expertise need not agree, data are noisy, questions may be too hard, and expertise differences may be too small to be relevant (e.g., Littlepage et al. 1995). This motivates a comparison of two strategies for group judgments: (i) deferring to the agent who is perceived as most competent, and (ii) taking the straight average of the estimates (Henry 1995; Soll and Larrick 2009). The overall outcomes suggest that the straight average is often surprisingly reliable, apparently being one of those "fast and frugal heuristics" (Gigerenzer and Goldstein 1996) that help boundedly rational agents to make cost-effective decisions.

On the other hand, even if not explicitly recognized as such, experts tend to exert greater influence on group judgments than non-experts (Bonner, Baumann and Dalal 2002). This motivates a principled epistemic analysis of the potential benefits of expertise-informed group judgments. We characterize conditions under which differentially weighted averages, fed by incomplete and perhaps distorted information on individual expertise, ameliorate group performance, compared to a straight average of the individual judgments. Our paper approaches this question from an analytical perspective, that is, with the help of a statistical model. We following the social permutation approach (e.g., Bonner 2000) and model the agents as unique entities with different abilities. This differs notably from more traditional social combination research where individual agents are modeled as interchangeable (e.g., Davis 1973). Our main result—that individual expertise makes a robust contribution to group performance—is not without surprise, given the generality of our conditions that also allow for perturbations such as individual bias or correlations among the group members. Therefore, our analytical results provide theoretical support to research on the recognition of experts in groups (e.g., Baumann and Bonner 2004), and they directly relate to empirical comparisons of differentially weighted group judgments to "composite judgments", such as the group mean or median (Einhorn, Hogarth and Klempner 1977; Hill 1982; Libby, Trotman and Zimmer 1987; Bonner 2004).

Our work is also related to two other research streams. First, there is a thriving epistemological literature on peer disagreement and rational consensus, where consensus is mostly reached by deference to (perceived) experts. However, this debate either focuses on social power and mutual respect relations (e.g., Lehrer and Wagner 1981), or on principled philosophical questions about resolving disagreement (e.g., Elga 2007). By means of a performance-focused mathematical model, we hope to bring this literature close to its primary target: the truth-tracking abilities of various epistemic strategies. There is also a vast literature on group decisions preference and judgment aggregation (e.g., List 2012), but two crucial features of our inquiry—the aggregation of numerical values and the particular role of experts—do not play a major role in there.

Second, there is a fast increasing body of literature on expert judgment and forecasting, which has emerged from applied mathematics and statistics and became a flourishing interdisciplinary field. This strand of research deals with the theoretical modeling of expert judgment, most notably the (Bayesian) reconciliation of probability distributions (Lindley 1983), but it also includes more practical questions such as comparison of calibration methods, choice of seed variables, analyses of the use of expert judgment in the past (Cooke 1991), and the study of general forecasting principles, such as the benefits of opinion diversity (Armstrong 2001; Page 2007). We differ from that approach in pooling individual (frequentist) estimators instead of subjective probability distributions, but we study similar phenomena, such as the impact of in-group correlations.

Admittedly, our baseline model is very simple, but due to this simplicity, we are able to prove a number of results regarding the behavior of differentially weighted estimates under correlation, bias and benchmark uncertainty. Here, our paper builds on analytical work in the forecasting and social psychology literature (Bates and Granger 1969; Hogarth 1978), following the approach of Einhorn, Hogarth and Klempner (1977).

The rest of the paper is structured as follows: we begin with explaining the model and stating conditions where differentially weighted estimates outperform the straight average (Sect. 2). In the sequel, we show that this relation is often preserved even if bias or mutual correlations are introduced (Sect. 3 and 4). Subsequently, we assess the impacts of over- and underconfidence (Sect. 5). Finally, we discuss our findings and wrap up our conclusions (Sect. 6).

2 The Model and Baseline Results

Our problem is to find a good estimate of an unknown quantity μ. For reasons of convenience, we assume without loss of generality that $\mu = 0$.[1]

[1] Rewriting our results for the general case $\mu \neq 0$ is just a matter of affine transformation, but comes with some notational baggage. Therefore we focus without loss of generality on $\mu = 0$.

We model the group members' individual estimates X_i, $i \leq n$, as independent random variables that scatter around the true value $\mu = 0$ with variance σ_i^2. The X_i are *unbiased* estimators of μ, that is, they have the property $\mathbb{E}[X_i] = \mu$. This baseline model is inspired by the idea that the agents try to approach the true value with a higher or lower degree of precision, but have no systematic bias in either direction. The competence of an agent is explicated as the degree of precision in estimating the true value. No further assumptions on the distributions of the X_i are made—only the first and second moments are fixed.

In this model, the question of whether the recognition of individual expertise is epistemically advantageous translates into the question of which convex combination of the X_i, $\hat{\mu} := \sum_{i=1}^{n} c_i X_i$, outperforms the straight average $\bar{\mu} := \frac{1}{n} \sum_{i=1}^{n} X_i$. Standardly, the quality of an estimate is assessed by its mean square error (MSE) which can be calculated as

$$\begin{aligned} \mathrm{MSE}(\hat{\mu}) := \mathbb{E}[(\hat{\mu} - \mu)^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i X_i\right)^2\right] \\ &= \sum_{i=1}^{n} c_i^2\, \mathbb{E}\left[X_i^2\right] + \sum_{i=1}^{n} \sum_{j \neq i} c_i c_j\, \mathbb{E}[X_i]\, \mathbb{E}[X_j] \\ &= \sum_{i=1}^{n} c_i^2\, \sigma_i^2 \qquad (1) \end{aligned}$$

which is minimized by the following assignment of the c_i (cf. Lehrer and Wagner 1981, 139):

$$c_i^* = \left(\sum_{j=1}^{n} \frac{\sigma_i^2}{\sigma_j^2} \right)^{-1}. \qquad (2)$$

Thus, naming the c_i^* as the "optimal weights" is motivated by two independent theoretical reasons:

1. As argued above, for independent and unbiased estimates X_i with variance σ_i^2, mean square error of the overall estimate is minimized by the convex combination $X = \sum_i c_i^* X_i$. Thus, for a standard loss function, the c_i^* are indeed the optimal weights.

2. Even when the square loss function is replaced by a more realistic alternative (Hartmann and Sprenger 2010), the c_i^* can still define the optimal convex combination of individual estimates. In that case, we

require stronger distributional assumptions.[2]

The problem with these optimal weights is that each agent's individual expertise would have to be known in order to calculate them. Given all the biases that actual deliberation is loaded with, e.g., ascription of expertise due to professional reputation, age or gender, or bandwagon effects, it is unlikely that the agents succeed at unraveling the expertise of all other group members (cf. Nadeau, Cloutier and Gray 1993; Armstrong 2001).

Therefore, we widen the scope of our inquiry:

> **Question:** Under which conditions will differentially weighted group judgments outperform the straight average?

A first answer is given by the following result where the differential weights preserve the expertise ranking:

Theorem 1 (First Baseline Result) *Let* $c_1, \ldots, c_n > 0$ *be the weights of the individual group members, that is,* $\sum_{i=1}^{n} c_i = 1$*. Without loss of generality, let* $c_1 \leq \ldots \leq c_n$*. Further assume that for all* $i > j$*:*

$$1 \leq \frac{c_i}{c_j} \leq \frac{c_i^*}{c_j^*} \tag{3}$$

Then the differentially weighted estimator $\hat{\mu} := \sum_{i=1}^{n} c_i X_i$ *outperforms the straight average. That is,* $MSE(\hat{\mu}) \leq MSE(\bar{\mu})$*, with equality if and only if* $c_i = 1/n$ *for all* $1 \leq i \leq n$*.*

This result demonstrates that relative accuracy, as measured by pairwise expertise ratios, is a good guiding principle for group judgments as long as the relative weights are not too extreme.

The following result extends this finding to a case where the benefits of differential weighting are harder to anticipate: we allow the c_i to lie in the entire $[1/n, c_i^*]$ (or $[c_i^*, 1/n]$) interval, allowing for cases where the ranking of the group members is not represented correctly. One might conjecture that this phenomenon adversely affects performance, but this is not the case:

Theorem 2 (Second Baseline Result) *Let* $c_1 \ldots c_n \in [0, 1]$ *such that* $\sum_{i=1}^{n} c_i = 1$*. In addition, let* $c_i \in [\frac{1}{n}; c_i^*]$ *respectively* $c_i \in [c_i^*; \frac{1}{n}]$ *hold for all* $1 \leq i \leq n$*. Then the differentially weighted estimator* $\hat{\mu} := \sum_{i=1}^{n} c_i X_i$ *outperforms the straight average. That is,* $\mathrm{MSE}(\hat{\mu}) \leq \mathrm{MSE}(\bar{\mu})$*, with equality if and only if* $c_i = 1/n$ *for all* $1 \leq i \leq n$*.*

[2] Hartmann and Sprenger (2010) prove the optimality of the c_i^* for the case of Normally distributed independent and unbiased estimates with variance σ_i^2 and the loss function family $L_\alpha(x) = 1 - \exp(-x^2/2\alpha^2)$. That paper also contains an elaborate justification for choosing this family of loss functions.

Note that none of the baseline results implies the other one. The conditions of the second result can be satisfied even when the ranking of the group members differs from their actual expertise, and a violation of the second condition (e.g., $c_i^* = 1/n$ and $c_i = 1/n + \varepsilon$) is compatible with satisfaction of the first condition. So the two results are really complementary.

We have thus shown that differential weighting outperforms straight averaging under quite general constraints on the individual weights, motivating the efforts to recognize experts in practice. The next sections extend these results to the presence of correlation and bias, thereby transferring them to more realistic circumstances.

3 Biased Agents

The first extension of our model concerns *biased* estimates X_i, that is, estimates that do not center around the true value $\mu = 0$, but around $B_i \neq 0$. We still assume that agents are honestly interested in getting close to the truth, but that training, experience, risk attitude or personality structure bias their estimates into a certain direction. For example, in assessing the impact of industrial development on a natural habitat, an environmentalist will usually come up with an estimate that significantly differs from the estimate submitted by an employee of an involved corporation—even if both are intellectually honest and share the same information.

For a biased agent i, the competence/precision parameter σ_i^2 has to be reinterpreted: it should be understood as the *coherence* (or non-randomness) of the agent's estimates instead of the accuracy. This value is indicative of accuracy only if the bias B_i is relatively small.

Under these circumstances, we can identify an intuitive sufficient condition for differential weighting to outperform straight averaging.

Theorem 3 *Let $X_1, \ldots, X_n$ be random variables with bias $B_1, \ldots, B_n$.*

(a) Suppose that the c_i in the estimator $\hat{\mu} = \sum_{i=1}^n c_i X_i$ satisfy one of the conditions of the baseline results (i.e., either $1 \leq c_i/c_j \leq c_i^*/c_j^*$ or $c_i \in [1/n, c_i^*]$ respectively $c_i \in [c_i^*, 1/n]$). In addition, let the following inequality hold:

$$\left(\sum_{i=1}^{n} c_i B_i\right)^2 < \left(\sum_{i=1}^{n} \frac{1}{n} B_i\right)^2 \tag{4}$$

Then differential weighting outperforms straight averaging, that is, $\mathrm{MSE}(\hat{\mu}) < \mathrm{MSE}(\bar{\mu})$.

(b) Suppose the following inequality holds:

$$\left(\sum_{i=1}^{n} c_i B_i\right)^2 > \left(\sum_{i=1}^{n} \frac{1}{n} B_i\right)^2 + \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 \qquad (5)$$

Then differential weighting does worse than straight averaging if condition (b) holds, that is, $\mathrm{MSE}(\hat{\mu}) > \mathrm{MSE}(\bar{\mu})$*.*

Intuitively, condition (4) states that the differentially weighted bias is smaller or equal than the average bias. As one would expect, this property favorably affects the performance of the differentially weighted estimator. Condition (5) states, on the other hand, that if the difference between the mean square biases of the weighted and the straight average exceeds the mean variance of the agents, then straight averaging performs better than weighted averaging.

When the group size grows to a very large number, both parts of Theorem 3 collapse into a single condition, as long as the biases and variances are both bounded. This is quite obvious since the second term of (5) is of the order $\mathcal{O}(1/n)$. Theorem 3 applies in particular in the case where agents are biased into the same direction and less biased agents make more coherent estimates (that is, with smaller variance):

Corollary 1 *Let $X_1, \ldots, X_n$, be random variables with bias $B_1, \ldots, B_n \geq 0$ such that $c_i \geq c_j$ implies $B_i \geq B_j$ (or vice versa for $B_1, \ldots, B_n \leq 0$). Then, with the same definitions as above:*

- $\mathrm{MSE}(\bar{\mu}) \geq \mathrm{MSE}(\hat{\mu})$.
- *If there is a uniform group bias, that is, $B := B_1 = \ldots = B_n$, then* $\mathrm{MSE}(\bar{\mu}) - \mathrm{MSE}(\hat{\mu})$ *is independent of B.*

So even if all agents have followed the same training, or have been raised in the same ideological framework, expertise recognition does not multiply that bias, but helps to increase the accuracy of the group's judgment. In particular, if there is a uniform bias in the group, the relative advantage of differential weighting is indepedent of the size of the bias. All in all, these results demonstrate the importance of expertise recognition even in groups where the members share a joint bias—a finding that is especially relevant for practice.

4 Independence Violations

We turn to violations of independence between the group members. Consider first the following fact that compares two groups with different degrees of correlation:

Fact 1 *If* $0 \leq \mathbb{E}[X_i X_j] \leq \mathbb{E}[Y_i Y_j]$ $\forall i \neq j \leq n$ *and* $\mathbb{E}[X_i^2] = \mathbb{E}[X_j^2]$*, then both straight averaging and weighted averaging on* X_i *yield a lower mean square error than the same procedures applied to* Y_i*.*

Fact 1 shows that less correlated groups perform better, ceteris paribus. For practical purposes, this suggests that heterogeneity of a group is an epistemic virtue since strong correlations between the agents are less likely to occur, making the overall result more accurate (cf. Page 2007).

Regarding the comparison of straight and weighted averaging, we can show the following result:

Theorem 4 *Let* $X_1, \ldots, X_n$ *be unbiased estimators, that is,* $\mathbb{E}[X_i] = \mu = 0$*, and let the* c_i *satisfy the conditions of one of the baseline results, with* $\hat{\mu}$ *defined as before. Let* $I \subseteq \{1, \ldots, n\}$ *be a subset of the group members with the property*

$$\forall i, j \neq k \in I : c_i \geq c_j \;\Rightarrow\; \mathbb{E}[X_j X_k] \geq \mathbb{E}[X_i X_k] \geq 0. \tag{6}$$

(i) **Correlation vs. Expertise** *If* $I = \{1, \ldots, n\}$*, then weighted averaging outperforms straight averaging, that is,* $\mathrm{MSE}(\hat{\mu}) \leq \mathrm{MSE}(\bar{\mu})$*.*

(ii) **Correlated Subgroup** *Assume that* $\mathbb{E}[X_i X_j] = 0$ *if* $i \in I$ *and* $j \notin I$*, and that*

$$\frac{1}{|I|} \sum_{i \in I} c_i \leq \frac{1}{n} \sum_{i=1}^{n} c_i. \tag{7}$$

Then weighted averaging still outperforms straight averaging, that is, $\mathrm{MSE}(\hat{\mu}) \leq \mathrm{MSE}(\bar{\mu})$*.*

To fully understand this theorem, we have to clarify the meaning of condition (6). Basically, it says that in group I, experts are less with correlated with other (sub)group members than non-experts.[3]

Once we have understood this condition, the rest is straightforward. Part (i) states that if I equals the entire group, then differential weighting

[3] Recall that $\mathbb{E}[X_i, X_k] \leq \mathbb{E}[X_j, X_k]$ can be rewritten as $\sigma_i/\sigma_j \leq \rho_{jk}/\rho_{ik}$ with ρ_{ij} defined as the Pearson correlation coefficient $\rho_{ij} := \mathbb{E}[X_i X_j]/\sigma_i \sigma_j$. Also, if $c_i \geq c_j$ then automatically $\sigma_i \leq \sigma_j$.

has an edge over averaging. That is, the benefits of expertise recognition are not offset by the perturbations that mutual dependencies may introduce. Arguably, the generality of the result is surprising since condition (6) is quite weak. Part (ii) states that differential weighting is also superior whenever there is no correlation with the rest of the group, and as long as the average competence in the subgroup is lower than the overall average competence (see equation (7)).

It is a popular opinion (e.g., Surowiecki 2004) that correlation of individual judgments is one of the greatest dangers for relying on experts in a group. To some extent, this opinion is vindicated by Fact 1 in our model. However, expertise-informed group judgments may still be superior to composite judgments, as demonstrated by Theorem 4. The interplay of correlation and expertise is subtle and not amenbale to broad-brush generalizations.

5 Over- and Underconfidence

We now consider a specific family of c_i's in order to study how group members' self-assessment in terms of quality affects group performance as a whole, modeled again as unbiased estimates X_i with variance σ_i^2.

Suppose that the group members have some idea of their own competence. That is, they are able to position themselves in relation to a commonly known *benchmark*: they are able to assess how much better or worse they expect themselves to perform compared to a default agent, modeled as a unbiased random variable with variance s^2. Such a scenario may be plausible when agents have a track record of their performance, or obtain performance feedback. The agents then express how much weight they should ideally get in a group of $n-1$ default agents:

$$c_i = \left(1 + \sum_{j \neq i} \frac{\sigma_i^2}{s^2}\right)^{-1} = \frac{s^2}{s^2 + (n-1)\sigma_i^2} \tag{8}$$

Assume further that every agent uses the same benchmark, that these weights also determine to what extent a group member compromises his or her own position, and that decision-making takes place on the basis of the normalized c_i's. It can then be shown (proof omitted) that the differentially weighted estimator $\hat{\mu}$ defined by equation (8) outperforms straight averaging—in fact, this is entailed by the Second Baseline Result (Theorem 2).

Here, we want to study how over- and underestimating the competence of a "default agent" will affect group performance. Is it always epistemically detrimental when the agents misguess the group competence?

The answer is, perhaps surprisingly, no. To explain this result, we first observe that the less confidence we have in the group (= s^2 is large), the more does the weighted average resemble the straight average. Recalling equation (8), we note that all c_i will be very close to 1. This implies that the expertise-informed average will roughly behave like the straight average.

Conversely, if the group is perceived as competent (=small value of s), then the c_i will typically *not* be close to 1 such that differential weights will diverge significantly from the straight average. This intuitive insight leads to the following theorem:

Theorem 5 *Let $\hat{\mu}_{s^2}$ and $\hat{\mu}_{\tilde{s}^2}$ be two weighted, expertise-informed estimates of μ, defined according to equation (8) with benchmarks s^2 and $\tilde{s}^2$, respectively. Then $MSE(\hat{\mu}_{s^2}) \leq MSE(\hat{\mu}_{\tilde{s}^2})$ if and only if $s^2 \leq \tilde{s}^2$.*

It can also be shown (proof omitted) that this procedure approximates the *optimal* weights c_i^* if the perceived group competence approaches perfection, that is, $s \to 0$. In other words, as long as the group members judge themselves accurately, optimism with regard to the abilities of the other group members is epistemically favorable. On the other hand, overconfidence in one's own abilities relative to the group typically deteriorates performance.

6 Discussion

We have set up an estimation model of group decision-making in order to study the effects of individual expertise on the quality of a group judgment. We have shown that in general, taking into account relative accuracy positively affects the epistemic performance of groups. Translated into our statistical model, this means that differential weighting outperforms straight averaging, even if the ranking of the experts is not represented accurately.

The result remains stable over several representative extensions of the model, such as various forms of bias, violations of independence, and over- and underconfident agents (Theorems 3–5). In particular, we demonstrated that differential weighting is superior (i) if experts are, on average, less biased; (ii) for a group of uniformly biased agents; (iii) if experts are less correlated with the rest of the group than other members. We also showed that uniform overconfidence in one's own abilities is detrimental for group performance whereas (over)confidence in the group may be beneficial. These properties may be surprising and demonstrate the stability and robustness of expertise-informed judgments, implying that the benefits of recognizing experts may offset the practical problems linked with that process.

Our model can in principle also be used for describing how groups actually form judgments. In that case, the involved tasks should neither be too intellective (that is, there is a *demonstrable* solution) or too judgmental (Laughlin and Ellis 1986): in highly intellective tasks, group will typically not perform better than the best individual (=the one who has solved the task correctly). This differs from our model where any agent has only partial knowledge of the truth. On the other hand, if the task is too judgmental, any epistemic component will be removed and the individual weights may actually be based on the *centrality* of a judgment, such as in Hinsz's (1999) SDS-Q scheme.

Finally, we name some distinctive traits of our model. First, unlike other models of group judgments that are detached from the group members' individual abilities (Davis 1973; DeGroot 1974; Lehrer and Wagner 1981; Hinsz 1999), it is a genuinely epistemic model, evaluating the performance of different ways of making a group judgment.[4] Thus, our model can be used normatively, for supporting the use of differential weights in group decisions, but also descriptively, for fitting the results of group decision processes.

Second, we did not make any specific distributional assumptions on how the agents estimate the target value. Our assumptions merely concern the first and second moment (bias and variance). We consider this parsimony a prudent choice because those distributions will greatly vary in practice, and we do not have epistemic access to them. Classical work in the social combination literature makes much more specific distributional assumptions (e.g., the multinomial distributions in Thomas and Fink 1961 and Davis 1973), restricting the scope of that analysis.

Third, we are not aware of other analytical models that take into account important confounders such as correlation, bias and over-/underconfident agents. Thus, we conclude that our model makes a substantial contribution to understanding the epistemic benefits of expertise in group judgments.

A Proofs of the Theorems

We will need the following inequalities repeatedly in the subsequent proofs. Let $c_1, \ldots, c_n > 0$. Then

$$\sum_{i=1}^{n} \frac{1}{c_i} \geq \frac{n^2}{\sum_{i=1}^{n} c_i} \tag{9}$$

[4] Lehrer and Wagner also defend their model from a normative point of view, but their arguments for this claim are not particularly persuasive, see e.g., Martini, Sprenger and Colyvan (2013).

with equality if and only if $c_1 = \ldots = c_n$. Moreover

$$n \sum_{i=1}^{n} c_i^2 \geq \left(\sum_{i=1}^{n} c_i \right)^2 \tag{10}$$

again with equality if and only if $c_1 = \ldots = c_n$. Both inequalities are special cases of the Power Mean Theorem (cf. Wilf 1985, 258).

For the First Baseline Result, we need the following

Lemma 1 *Let $k < n$ and let $(c_1, \ldots, c_n)$ be a sequence such that*

(1) $\sum_{i=1}^{n} c_i = s$ for some $s > 0$ and all c_i are positive;

(2) $c_1 = \ldots = c_k$ and $c_{k+1} = \ldots = c_n$;

(3) $c_k \leq c_{k+1}$ and $1 \leq \frac{c_{k+1}}{c_k} \leq \frac{c^*_{k+1}}{c^*_k}$.

Further assume that $\sigma_1 \geq \ldots \geq \sigma_n$. Then

$$\sum_{i=1}^{n} \left(\frac{s}{n} \right)^2 \sigma_i \geq \sum_{i=1}^{n} c_i^2 \sigma_i$$

Furthermore, we show that under the above conditions (i.e. $\sum_{i=1}^{n} c_i = s$), the value of the sum $\sum_{i=1}^{n} c_i^2 \sigma_i$ decreases as the quotient $\frac{c_{k+1}}{c_k}$ increases.

Proof of Lemma 1: Fix r such that

- $c_i = \frac{s}{n} - \frac{r}{k}$ for $i \leq k$
- $c_i = \frac{s}{n} + \frac{r}{n-k}$ for $i > k$

Then we have to show that:

$$\sum_{i \leq k} \left(\frac{s}{n} - \frac{r}{k} \right)^2 \sigma_i + \sum_{i > k} \left(\frac{s}{n} + \frac{r}{n-k} \right)^2 \sigma_i - \sum_{i=1}^{n} \left(\frac{s}{n} \right)^2 \sigma_i \leq 0$$

The above equation reduces to:

$$r^2 \left(\sum_{i \leq k} \frac{1}{k^2} \sigma_i + \sum_{i > k} \frac{1}{(n-k)^2} \sigma_i \right) - \frac{2s}{n} r \left(\sum_{i \leq k} \frac{1}{k} \sigma_i - \sum_{i > k} \frac{1}{n-k} \sigma_i \right) \leq 0 \tag{11}$$

Now the left hand side of the above equation is a quadratic function in r with zeros at 0 and

$$r_0 = \frac{2s}{n} \frac{\sum_{i\le k} \frac{1}{k}\sigma_i - \sum_{i>k} \frac{1}{n-k}\sigma_i}{\sum_{i\le k} \frac{1}{k^2}\sigma_i + \sum_{i>k} \frac{1}{(n-k)^2}\sigma_i} \tag{12}$$

Since the σ_i are ordered decreasingly we get

$$r_0 \ge \frac{2s}{n} \frac{\sum_{i\le k} \frac{1}{k}\sigma_i - \sigma_{k+1}}{\sum_{i\le k} \frac{1}{k^2}\sigma_i + \frac{1}{(n-k)}\sigma_{k+1}}$$

Now this is a function of the form $\frac{kx-a}{x+b}$ with $a, b > 0$. Since these functions are increasing for $x > -b$, the inequality above can be strengthened to

$$r_0 \ge \frac{2s}{n} \frac{\sigma_k - \sigma_{k+1}}{\frac{1}{k}\sigma_k + \frac{1}{(n-k)}\sigma_{k+1}}$$

Recall that $\frac{c_{k+1}}{c_k} \le \frac{c^*_{k+1}}{c^*_k} = \frac{\sigma_k}{\sigma_{k+1}} =: \sigma$ Inserting this transforms the above equation into:

$$r_0 \ge \frac{2s}{n} \frac{(\sigma - 1)\sigma_{k+1} k(n-k)}{\sigma_{k+1}((n-k)\sigma + k)}$$

Our assumptions about the c_i translate into

$$\frac{\frac{s}{n} + \frac{r}{n-k}}{\frac{s}{n} - \frac{r}{k}} \le \frac{c^*_{k+1}}{c^*_k} = \frac{\sigma_k}{\sigma_{k+1}}$$

This transforms to

$$r \le \frac{s}{n} \frac{(\sigma - 1)k(n-k)}{(n-k) - \sigma k}$$

In particular $r < r_0$, finishing the proof of (11). For the last statement of Lemma 1, observe that the left hand side of (11) is a quadratic function with minimum $\frac{1}{2}r_0$, and that $r \le \frac{1}{2}r_0$. □

Proof of Theorem 1: By assumption the c_i are ordered increasingly, thus the σ_i are ordered decreasingly. For a vector of weights $\mathbf{w} \in \mathbb{R}^n$ (i.e. all w_i positiv and $\sum_i w_i = 1$), we denote the mean square error of the estimator $\sum w_i X_i$ by $\Psi(\mathbf{w})$: That is:

$$\Psi(\mathbf{w}) := \sum w_i^2 \sigma_i$$

Thus for $\mathbf{c} = (c_1 \dots c_n)$ as in the theorem we have to show $\Psi(\mathbf{c}) \le \Psi(\mathbf{e})$, where $\mathbf{e}$ is the equal weight vector $(\frac{1}{n}, \dots, \frac{1}{n})$. To this end we will construct a sequence of weight vectors $\mathbf{e} = \mathbf{d}_0, \dots, \mathbf{d}_{n-1} = \mathbf{c}$ such that:

(i) each $\mathbf{d}_i$ satisfies the assumptions of Theorem 1;

(ii) for $\mathbf{d}_i = (d_1 \dots d_n)$, there is some $k \in \mathbb{N}$ such that

$$d_1 = \dots = d_k \text{ and } d_1 > c_1; \dots; d_k > c_k;$$

$$d_j = c_j \text{ for } k < j \leq k+i \qquad \text{(where } i \text{ is the index of } \mathbf{d}_i\text{);}$$

$$d_{k+i+1} = \dots = d_n \text{ and } d_{k+i+1} \leq c_{k+i+1}; \dots; d_n \leq c_n;$$

(iii) $\Psi(d_{i-1}) \geq \Psi(d_i)$.

Thus $\mathbf{d}_{i-1} = \mathbf{c}$ and $\Psi(\mathbf{c}) \leq \Psi(\mathbf{e})$ as desired. The $\mathbf{d}_i$ are constructed inductively as follows: Assume $\mathbf{d}_{i-1} = (d'_1 \dots d'_n)$ has already been constructed. If $i = 1$ let k be the unique index such that $c_k < \frac{1}{n}$ and $c_{k+1} \geq \frac{1}{n}$. If $i > 1$ let k be as in the above conditions for $\mathbf{d}_{i-1}$. First note that if $k = 0$, then $d'_j \leq c_j$ for all j and thus $\mathbf{d}_{i-1} = \mathbf{c}$ since both are weight vectors and we are done. Thus assume $k \geq 1$ for the rest of the proof. With a similar argument, we can show that $k + i + 1 \leq n$. Now choose the maximal $r \in \mathbb{R}$ that satisfies

$$d'_k - c_k \geq \frac{r}{k} \qquad\qquad c_{k+i+1} - d'_{k+i+1} \geq \frac{r}{n-k-i-1} \tag{13}$$

By the above conditions, $r \geq 0$. Then define $\mathbf{d}_i = (d_1, \dots, d_n)$ by:

- $d_j = d'_j - \frac{r}{k}$ for $j \leq k$;
- $d_j = c_j$ for $k < j \leq k+i$;
- $d_j = d'_j + \frac{r}{n-k-i-1}$ for $j \geq k+i+1$.

To see that $\mathbf{d}_i$ satisfies conditions (i)-(iii), first note that since r was chosen to be maximal, one of the two inequalities in (13) has to be an equality. Thus we either have $d_k = c_k$ or $d_{k+i+1} = c_{k+i+1}$ and condition (ii) is satisfied. Further note that

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} d'_i - \sum_{i \leq k} + \frac{r}{k} + \sum_{i \geq k+i+1} \frac{r}{n-k-i-1} = 1$$

Using that the c_i are ordered increasingly, it is easy to see that $\mathbf{d}_i$ satisfies the assumptions of Theorem 1. Furthermore, applying the monotonicity part of Lemma 1 to the set of indices $I := \{1, \dots, k\} \cup \{i+k+1, \dots, n\}$, we get $\sum_I d_i \sigma_i^2 \leq \sum_I d'_i \sigma_i^2$. Thus $\Psi(\mathbf{d}_i) \leq \Psi(\mathbf{d}_{i-1})$ since $\mathbf{d}_{i-1}$ and $\mathbf{d}_i$ coincide outside I. This finishes the proof. □

Proof of Theorem 2: We would like to show that the mean square error of the straight average $\bar{\mu} := (1/n)\sum_{i=1}^{n} X_i$ exceeds the mean square error of the weighted estimate $\hat{\mu}$. The MSE difference can be calculated as

$$\begin{aligned}\Delta(c_1,\ldots,c_n) &:= \text{MSE}(\bar{\mu}) - \text{MSE}(\hat{\mu}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 - \sum_{i=1}^{n} c_i^2\sigma_i^2 \\ &= \frac{1}{n^2}\left(\sum_{j=1}^{n}\frac{1}{\sigma_j^2}\right)^{-1}\sum_{i=1}^{n}\frac{1}{c_i^*}\left(1-n^2c_i^2\right)\end{aligned}$$

where we have made use of $\mathbb{E}[X_i X_j] = 0,\ \forall i \neq j$, and of $c_i^* = \left(\sum_{j=1}^{n}\frac{\sigma_i^2}{\sigma_j^2}\right)^{-1}$ (cf. equation (2)). Thus, instead of considering Δ, it suffices to show that

$$\Delta'(c_1 \ldots c_n) := \sum_{i=1}^{n}\frac{1}{c_i^*}\left(1-n^2c_i^2\right) \geq 0.$$

To this end, let $I_i := [1/n; c_i^*]$ (respectively $[c_i^*; 1/n]$) and let $\mathcal{Q} := I_1 \times \ldots \times I_n$. Then,

$$\mathcal{D} := \mathcal{Q} \cap \{(c_1,\ldots,c_n) | \sum_{i=1}^{n} c_i = 1\}$$

defines the "domain" of our theorem, and it is a polygon. Moreover, since $\sum_i \frac{n^2}{c_i^*}c_i^2$ is a positive determinate quadratic form in the c_i, we get that $\Delta'^{-1}([0;\infty))$ is convex. Thus, it suffices to show that Δ' is positive on the vertices of $\mathcal{D}$. Note that since $\{x | \sum x_i = 1\}$ is of dimension $n-1$, the vertices of $\mathcal{D}$ are of the form $\mathbf{v} = (c_1^*,\ldots,c_{k-1}^*, c_k, 1/n,\ldots,1/n)$—the ordering is assumed for convenience, and c_k is defined such that $||\mathbf{v}||_1 = 1$. Thus we have to show that $\Delta'(c_1^*,\ldots,c_{k-1}^*, c_k, 1/n,\ldots,1/n) \geq 0$.

In the case $k = 1$, the desired inequality holds trivially since $c_k = 1 - (n-1)\cdot(1/n) = 1/n$. Thus we assume $k > 1$ for the remainder of this proof. Let l denote the real number satisfying

$$\sum_{i=1}^{n} c_i^* = l\frac{k-1}{n}$$

Observe that for $c_i = \frac{1}{n}$ the corresponding summands in Δ' vanish. Thus we have to show that

$$\sum_{i=1}^{k-1}\frac{1}{c_i^*}\left(1-n^2c_i^{*2}\right) + \frac{1}{c_k^*}\left(1-n^2c_k^2\right) \geq 0$$

Using the definition of l from above and inequality (9) gives $\sum_{i=1}^{k-1} \frac{1}{c_i^*} \geq (k-1)^2/(\sum_{i=1}^{k-1} c_i) \geq \frac{n(k-1)}{l}$. Thus, it suffices to show

$$n(k-1)\left(\frac{1}{l} - l\right) + \frac{1}{c_k^*}\left(1 - n^2 c_k^2\right) \geq 0 \tag{14}$$

Since the c_i add up to one, we can express the dependency between l and c_k by

$$c_k = \frac{(k-1)(1-l)+1}{n} \qquad \text{or by} \qquad l = \frac{k - nc_k}{k-1} \tag{15}$$

Inserting this into (14) gives

$$\begin{aligned}
\Delta'(c_1, \ldots, c_n) &= \left(\frac{1}{l} - l\right) n(k-1) - \frac{1}{c_k^*}\left((1-l)^2(k-1)^2 + 2(1-l)(k-1)\right) \\
&= \frac{k-1}{l}\left[\left(1 - l^2\right) n - \frac{l}{c_k^*}\left((1-l)^2(k-1) + 2(1-l)\right)\right] \\
&= \frac{k-1}{l}\left[(1-l)\left((1+l)n - \frac{l}{c_k^*}((1-l)(k-1)+2)\right)\right]
\end{aligned}$$

Since the first factor is always positive, it suffices to show that the factor in the square brackets, denoted by $P(l)$, is positive for every l that can occur in our setting. We do this by a case distinction on the value of c_k^*

Case 1: $c_k^* \leq 1/n$. Noting $c_k \in [c_k^*, \frac{1}{n}]$ and the dependency (15) between l and c_k, we have to show that $P(l) \geq 0$ for all $l \in [1; \frac{k - nc_k^*}{k-1}]$. We observe that P is a polynomial of third order with zero points of P given by $P(1) = 0$ and

$$r_{\pm} = \frac{k + 1 - nc_k^* \pm \sqrt{(k+1-nc_k^*)^2 - 4(k-1)c_k^* n}}{2(k-1)}$$

with r_+ denoting the larger of these two numbers. With some algebra it also follows that $P'(1) \geq 0$ if and only if $c_k* \leq 1/n$. From the functional form of $P(l)$—a polynomial of the third degree with negative leading coefficient—we can then infer that $l = 1$ must be the middle zero point of P. To prove that $P(l) \geq 0$ in the critical interval, it remains to show that for the rightmost zero point, we have $r_+ \geq \frac{k - nc_k^*}{k-1}$:

$$\begin{aligned}
\frac{k - nc_k^*}{k-1} &\leq r_+ \\
\Leftrightarrow \frac{2k - 2nc_k^*}{2(k-1)} &\leq \frac{k + 1 - c_k^* n + \sqrt{(k+1-nc_k^*)^2 - 4(k-1)c_k^* n}}{2(k-1)} \\
\Leftrightarrow k - 1 - nc_k^* &\leq \sqrt{(k+1-nc_k^*)^2 - 4(k-1)c_k^* n} \\
\Leftrightarrow c_k^* n &\leq 1
\end{aligned}$$

completing the proof for the case $c_k* \leq 1/n$.

Case 2: $c_k* \geq 1/n$. In this case we are dealing with the interval $l \in [\frac{k-nc_k^*}{k-1}; 1]$. The same calculations as above yield

$$\frac{k - nc_k^*}{k-1} \geq r_+ \qquad \text{if and only if} \qquad c_k^* n \geq 1.$$

in particular $r_+ < 1$. Thus l always lies between the middle and the right-most zero point of $P(l)$, and in particular, $P(l) \geq 0$ for all $l \in [\frac{k-nc_k^*}{k-1}; 1]$. □

Proof of Theorem 3: Let the X_i center around $B_i > 0$. Then $\mathbb{E}[X_i - B_i] = 0$, and we observe

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - B_i)\right)^2\right] + \left(\frac{1}{n}\sum_{i=1}^{n} B_i\right)^2$$

Analogously, we obtain

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} c_i X_i\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i(X_i - B_i)\right)^2\right] + \left(\sum_{i=1}^{n} c_i B_i\right)^2.$$

Like in Theorem 2, we define $\Delta(c_1, \ldots, c_n) := \text{MSE}(\bar{\mu}) - \text{MSE}(\hat{\mu})$ as the difference in mean square error between both estimates and show that $\Delta(c_1, \ldots, c_n) \geq 0$ if equation (4) is satisfied.

$$\begin{aligned}\Delta(c_1, \ldots, c_n) &:= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - B_i)\right)^2\right] - \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i(X_i - B_i)\right)^2\right] \\ &+ \left(\frac{1}{n}\sum_{i=1}^{n} B_i\right)^2 - \left(\sum_{i=1}^{n} c_i B_i\right)^2 \qquad (16)\end{aligned}$$

By Theorem 1 and/or Theorem 2, the first line is greater or equal to zero, and by equation (4), the second line is also non-negative. Thus $\Delta(c_1, \ldots, c_n) \geq 0$, showing the superiority of differential weighting.
For the second part of the theorem, we just observe that

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - B_i)\right)^2\right] - \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i(X_i - B_i)\right)^2\right] \geq \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2.$$

□

Proof of Corollary 1: It is easy to see that the conditions of the corollary satisfy the requirements of part (a) of Theorem 3. This yields the desired result for the first part of the theorem. For the second, part, let the X_i all center around $B \neq 0$. Then $X_i - B$ is unbiased, and we observe

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - B)\right)^2\right] + B^2 \\
\mathbb{E}\left[\left(\sum_{i=1}^{n} c_i X_i\right)^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i (X_i - B)\right)^2\right] + B^2.
\end{aligned}$$

Therefore, under the conditions of the theorem,

$$\Delta(c_1, \ldots, c_n) = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - B)\right)^2\right] - \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i (X_i - B)\right)^2\right]$$

showing that Δ only depends on the centered estimates. □

Proof of Fact 1: First we deal with straight averaging:

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] - \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}\left[X_i X_j\right] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}\left[Y_i Y_j\right] \geq 0$$

The proof exploits that X_i and Y_i have the same variance, thus $\mathbb{E}\left[X_i^2\right] = \mathbb{E}\left[Y_i^2\right]$. The proof for differential weights is similar, making use of the fact that the c_i are the same for X_i and Y_i because they only depend on the variance of the random variable. □

Proof of Theorem 4, part (i): First, assume without loss of generality that $c_i \geq c_{i+1}$ for all $i < n$. Thus, our assumption on the $\mathbb{E}[X_i X_j]$ reduces to $\mathbb{E}[X_i X_k] \leq \mathbb{E}[X_j X_k]$ for $i \geq j \neq k$. First, we show the theorem under the assumption that all $\mathbb{E}[X_i X_j]$ with $i \neq j$ are equal, say $\mathbb{E}[X_i X_j] = \gamma$. By Theorem 1 and/or 2, it suffices to show that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}\left[X_i, X_j\right] - \sum_{i=1}^{n}\sum_{j\neq i} c_i c_j\,\mathbb{E}\left[X_i X_j\right] \geq 0$$

Inserting $\mathbb{E}[X_iX_j] = \gamma$ this reduces to

$$\gamma \cdot \left(\frac{n-1}{n} - \sum_{i=1}^{n} \sum_{j \neq i} c_i\, c_j \right) \geq 0 \tag{17}$$

The point $(1/n, \ldots, 1/n)$ is a global minimum of the function $f(\mathbf{x}) = \sum_i x_i^2$ under the constraints $x_1, \ldots, x_n \geq 0$ and $\sum_i x_i = 1$. Thus we have

$$\frac{1}{n} = f\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \leq f(\mathbf{c}) = \sum_{i=1}^{n} {c_i}^2 \tag{18}$$

Observing $\sum_{i=1}^n \sum_{j=1}^n c_i\, c_j = \left(\sum_{i=1}^n c_i\right)^2 = 1$ and combining this equality with (17) and (18), we obtain

$$\frac{n-1}{n} - \sum_{i=1}^{n} \sum_{j \neq i} c_i\, c_j \quad = \quad \frac{n-1}{n} - \sum_{i=1}^{n} \sum_{j=1}^{n} c_i\, c_j + \sum_{i=1}^{n} {c_i}^2 \geq 0 \tag{19}$$

thus proving the statement in the case that all $\mathbb{E}[X_iX_j]$ are the same.

For the general case let us assume that not all c_i are the same (otherwise the theorem is trivially true). Thus we either have $c_1 > c_{n-1}$ or $c_2 > c_n$ since the c_i are ordered decreasingly. In the following, we assume $c_2 > c_n$, the other case works with a similar argument. First observe that

$$\sum_{i=1}^{n} \sum_{j \neq i} c_i c_j\, \mathbb{E}\left[X_iX_j\right] = 2 \sum_{i=1}^{n} \sum_{j<i} c_i c_j\, \mathbb{E}\left[X_iX_j\right].$$

Thus, we can concentrate on $\{\mathbb{E}[X_iX_j] | i > j\}$. We fix a natural number c and let S_c be the set of all vectors $(\mathbb{E}[X_iX_j])_{(i>j)}$ fulfilling the conditions of our theorem and $\sum_{i>j} \mathbb{E}[X_iX_j] = c$ We then consider the functional

$$\begin{aligned} \tilde{\varphi}(e) &:= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j<i} \mathbb{E}\left[X_iX_j\right] - \sum_{i=1}^{n} \sum_{j<i} c_i c_j\, \mathbb{E}\left[X_iX_j\right] \\ &= \frac{1}{2} \left[\sum_{i=1}^{n} \sum_{j \neq i} \mathbb{E}\left[X_iX_j\right] - \sum_{i=1}^{n} \sum_{j \neq i} c_i c_j\, \mathbb{E}\left[X_iX_j\right] \right] \end{aligned}$$

on S_c. Observe that every S_c contains exactly one point e_{eq} where all $\mathbb{E}[X_iX_j]$ are equal. By the first part of this proof, $\tilde{\varphi}(e_{eq})$ is non-negative. Thus, it suffices to show that e_{eq} is an absolute minimum of $\tilde{\varphi}$ on S_c. First, observe that the value of $\frac{1}{n^2} \sum_{i=1}^n \sum_{j<i} \mathbb{E}\left[X_i, X_j\right]$ is constantly $\frac{c}{n^2}$ on S_c, thus it suffices to show that

$$\varphi(e) := \sum_{i=1}^{n} \sum_{j<i} c_i c_j\, \mathbb{E}\left[X_iX_j\right] \tag{20}$$

attains its maximum on S_c in e_{eq}.

To do so, we show the following: For every $e \in S_c$ with $e \neq e_{eq}$ there is some $e' \in S_c$ with $\varphi(e') > \varphi(e)$. In particular, φ does not take its maximum on S_c in e. Thus assume that $e = (\mathbb{E}[X_i X_j])_{(i>j)} \in S_c$ is given. Since $e \neq e_{eq}$ there are some indices $s > t$ and $k > l$ such that $\mathbb{E}[X_s X_t] \neq \mathbb{E}[X_k X_l]$. Furthermore, we can assume that $t \geq l$. Without loss of generality (by potentially replacing one of the two entries with $\mathbb{E}[X_s X_l]$) we can assume that either $s = k$ or $t = l$. In the following we assume $s = k$, the other case works similar. The idea of the following construction is: We show that moving towards a more equal distribution of the entries $\mathbb{E}[X_i X_j]$ increases $\varphi(e)$. In particular, we construct $e' = (\mathbb{E}'[X_i X_j])_{(i>j)} \in S_c$ as follows: In every row $r_i := \langle \mathbb{E}[X_i X_1] \dots \mathbb{E}[X_i X_{i-1}] \rangle$ of e we replace all the entries of this row by their arithmetic mean. Formally, that is for all i and j (independent of j):

$$\mathbb{E}'[X_i X_j] = \frac{1}{i-1} \sum_{l<i} \mathbb{E}[X_i X_l]$$

Trivially this operation satisfies for all i:

$$\sum_{j<i} \mathbb{E}[X_i X_j] = \sum_{j<i} \frac{1}{i-1} \sum_{j<i} \mathbb{E}[X_i X_j] = \sum_{j=1}^{i-1} \mathbb{E}'[X_i X_j]$$

and thus also for the double sum:

$$\sum_{i=1}^{n} \sum_{j<i} \mathbb{E}[X_i X_j] = \sum_{i=1}^{n} \sum_{j<i} \mathbb{E}'[X_i X_j].$$

In particular e' is in S_c. Furthermore, we have assumed that the c_i are ordered decreasingly. Recall that $c_k > c_j$ implies $\mathbb{E}[X_i X_k] \leq \mathbb{E}[X_i X_j]$ by assumption, therefore the rows r_i were ordered increasingly, and thus the rows of $e' - e$:

$$\mathbb{E}'[X_i, X_1] - \mathbb{E}[X_i X_1]; \dots ; \mathbb{E}'[X_i, X_{i-1}] - \mathbb{E}[X_i X_{i-1}]$$

are ordered decreasingly (since the rows of e' are constant). In particular, we have for any i:

$$0 = \sum_{j<i} \mathbb{E}'[X_i X_j] - \mathbb{E}[X_i X_j] \leq \sum_{j<i} c_i c_j (\mathbb{E}'[X_i X_j] - \mathbb{E}[X_i X_j]) \tag{21}$$

where the $\leq$ comes from the fact that both c_j and $\mathbb{E}'[X_i X_j] - \mathbb{E}[X_i X_j]$ are decreasing in j. Summing that up over all i we get that

$$0 = \sum_{i=1}^{n} \sum_{j<i} \mathbb{E}'[X_i X_j] - \mathbb{E}[X_i X_j] \leq \sum_{i=1}^{n} \sum_{j<i} c_i c_j \left(\mathbb{E}'[X_i X_j] - \mathbb{E}[X_i X_j] \right) = \varphi(e') - \varphi(e)$$

Thus we have $\varphi(e') \geq \varphi(e)$ as desired. Now observe that (21) for $i = s$ is the following:

$$\begin{aligned} 0 &= \sum_{j<s} \mathbb{E}'[X_s X_j] - \mathbb{E}[X_s X_j] \\ &= \sum_{j<s, j\neq t,l} \big(\mathbb{E}'[X_s X_j] - \mathbb{E}[X_s X_j]\big) + \mathbb{E}'[X_s X_t] - \mathbb{E}[X_s X_t] + \mathbb{E}'[X_s X_l] - \mathbb{E}[X_s X_l] \end{aligned}$$

with both,

$$\sum_{j<s, j\neq t,l} \mathbb{E}'[X_s X_j] - \mathbb{E}[X_s X_j] \leq \sum_{j<s, j\neq t,l} c_s c_j \left(\mathbb{E}'[X_s X_j] - \mathbb{E}[X_s X_j]\right)$$

and

$$\begin{aligned} &\mathbb{E}'[X_s X_t] - \mathbb{E}[X_s X_t] + \mathbb{E}'[X_s X_l] - \mathbb{E}[X_s X_l] \\ &\leq c_s c_t (\mathbb{E}'[X_s X_t] - \mathbb{E}[X_s X_t]) + c_s c_l (\mathbb{E}'[X_s X_l] - \mathbb{E}[X_s X_l]). \end{aligned}$$

By construction we have $\mathbb{E}[X_s X_t] \neq \mathbb{E}[X_s X_l]$, thus we would have a strict inequality in the last summand (and thus in the entire sum) if we knew that $c_t \neq c_l$. Unfortunately, this is not always the case. However, we have put ourselves in a situation where applying the same construction again with $\mathbb{E}'[X_2 X_1]$ and $\mathbb{E}'[X_n X_1]$ replacing $\mathbb{E}[X_s X_t]$ and $\mathbb{E}[X_s X_l]$ yields the desired (since we have assumed that $c_2 > c_n$. To see this, observe that

- $\mathbb{E}[X_2 X_1] = \mathbb{E}'[X_2 X_1]$ by construction
- $\mathbb{E}'[X_s X_1] > \mathbb{E}[X_s X_1]$ since $\mathbb{E}[X_s X_t] \neq \mathbb{E}[X_s, X_l]$ and $\mathbb{E}[X_s X_1]$ is the minimal element in the row r_s
- $\mathbb{E}[X_2 X_1] \leq \mathbb{E}[X_s X_1]$ by assumption

Thus we have

$$\mathbb{E}'[X_2 X_1] = \mathbb{E}[X_2 X_1] \leq \mathbb{E}[X_s X_1] < \mathbb{E}'[X_s X_1] \leq \mathbb{E}'[X_n X_1]$$

By assumption we have $c_2 > c_n$ and repeating the construction from above with coloumns replacing rows and $\mathbb{E}'[X_2, X_1], \mathbb{E}'[X_n, X_1]$ as the two reference points yields the desired.

Proof of Theorem 4, part (ii): We have to show that the statement holds if all $\mathbb{E}[X_i X_j]$ with $i \neq j \in I$ are the same. The step from this case to the general statement works as in the proof above. As in the proof of *i*), it suffices to show that

$$\frac{1}{n^2}\sum_{i\in I}\sum_{j\neq i\in I} 1 \geq \sum_{i\in I}\sum_{j\neq i\in I} c_i c_j$$

Let $\bar{c} = \frac{1}{|I|}\sum_{i\in I} c_i$. By equation (10) we have

$$\sum_{i\in I} c_i^2 \geq \frac{1}{|I|}\left(\sum_{i\in I} c_i\right)^2 = \frac{1}{|I|}|I|^2\bar{c}^2 = |I|\bar{c}^2$$

thus

$$\sum_{i\in I}\sum_{j\neq i\in I} c_i c_j \leq (|I|^2 - |I|)\bar{c}^2 \leq |I|^2 - |I| = \frac{1}{n^2}\sum_{i\in I}\sum_{j\neq i\in I} 1$$

with the last inequality coming from our assumption that $\bar{c} < 1$.

Proof of Theorem 5: Let the benchmark agent have standard deviation $s > 0$, that is, variance s^2. We will show that $\Delta(s, \sigma_1, \ldots, \sigma_n)$—the MSE difference between the differentially weighted and the straight average—is strictly monotonically decreasing in the first argument. To this effect, we calculate

$$\Delta(s, \sigma_1, \ldots, \sigma_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 - \left(\frac{1}{\sum_k c_k}\right)^2 \sum_{i=1}^{n} c_i^2\sigma_i^2.$$

Now we show that $\frac{\partial}{\partial s}\Delta(s, \sigma_1, \ldots, \sigma_n) \leq 0$, where c_i' denotes $(\partial/\partial)sc_i$:

$$\begin{aligned}
\frac{\partial}{\partial s}\Delta(s, \sigma_1, \ldots, \sigma_n) &= -\frac{\partial}{\partial s}\left(\sum_{i=1}^{n}\frac{c_i^2}{(\sum_k c_k)^2}\sigma_i^2\right) \\
&= -\sum_{i=1}^{n}\sigma_i^2 \cdot 2 \cdot \left(\frac{c_i}{\sum_k c_k}\right)\frac{c_i'\sum_j c_j - c_i\sum_j c_j'}{(\sum_k c_k)^2} \\
&= -\frac{2}{(\sum_k c_k)^3}\sum_{i=1}^{n}\sigma_i^2 c_i\left(\sum_{j\neq i} c_i' c_j - c_i c_j'\right) \\
&= -\frac{2}{(\sum_k c_k)^3}\sum_{i=1}^{n}\sum_{j<i}\left(\sigma_i^2 c_i - \sigma_j^2 c_j\right)\left(c_i' c_j - c_i c_j'\right)
\end{aligned}$$

Since we are only interested in the sign of the first derivative and $-\frac{2}{(\sum_k c_k)^3} < 0$, it suffices to show that:

$$\left(\sigma_i^2 c_i - \sigma_j^2 c_j\right)\left(c_i' c_j - c_j' c_i\right) \geq 0 \qquad (22)$$

We show that the terms in both brackets have the same sign.
For the first bracket we have:

$$\begin{aligned}\sigma_i^2 c_i - \sigma_j^2 c_j &= s^2 \frac{\sigma_i^2}{s^2 + (n-1)\sigma_i^2} - s^2 \frac{\sigma_j^2}{s^2 + (n-1)\sigma_j^2} \\ &= s^4 \frac{\sigma_i^2 - \sigma_j^2}{(s^2 + (n-1)\sigma_i^2)(s^2 + (n-1)\sigma_j^2)}\end{aligned}$$

which is larger than or equal to 0 if and only if $\sigma_i^2 > \sigma_j^2$. Similarly, we observe for the second bracket that

$$c_i' = \frac{2(n-1)s\sigma_i^2}{(s^2 + (n-1)\sigma_i^2)^2}.$$

which allows us to conclude

$$\begin{aligned}& c_i' c_j - c_j' c_i \\ = \; & \frac{2(n-1)s\sigma_i^2}{(s^2 + (n-1)\sigma_i^2)^2} \cdot \frac{s^2}{s^2 + (n-1)\sigma_j^2} - \frac{2(n-1)s\sigma_j^2}{(s^2 + (n-1)\sigma_j^2)^2} \cdot \frac{s^2}{s^2 + (n-1)\sigma_i^2} \\ = \; & 2(n-1)s^5 \; \frac{\sigma_i^2 - \sigma_j^2}{(s^2 + (n-1)\sigma_i^2)^2 (s^2 + (n-1)\sigma_j^2)^2}\end{aligned}$$

Thus, both factors in (22) have the same sign, implying $\frac{\partial}{\partial s}\Delta(s, \sigma_1, \ldots, \sigma_n) \leq 0$ which is want we wanted to prove. □

References

Armstrong, J. Scott (2001): "Combining Forecasts", in: J. Scott Armstrong (eds.), *Principles of Forecasting: A Handbook For Researchers and Practitioners.* Norwell, MA: Kluwer Academic Publishers.

Bates, J.M., and C.W.J. Granger (1969): "The combination of forecasts", *Operational Research Quarterly* 20, 451–468.

Baumann, Michael R., and Bryan L. Bonner (2004): "The effects of variability and expectations on utilization of member expertise and group performance", *Organizational Behavior and Human Decision Processes* 93, 89–101.

Bonner, Bryan L. (2000): "The effects of extroversion on influence in ambiguous group tasks", *Small Group Research* 31, 225–244.

Bonner, Bryan L. (2004): "Expertise in Group Problem Solving: Recognition, Social Combination, and Performance", *Group Dynamics: Theory, Research, and Practice* 8, 277–290.

Bonner, Bryan L., Michael R. Baumann, and Reeshad S. Dalal (2002): "The effects of member expertise on group decision-making and performance", *Organizational Behavior and Human Decision Processes* 88, 719–736.

Clemen, Robert T. (1989): "Combining forecasts: a review and annotated bibliography", *International Journal of Forecasting* 5, 559–583.

Cooke, Roger M. (1991): *Experts in uncertainty.* Oxford: Oxford University Press.

Davis, James H. (1973): "Group decision and social interaction: A theory of social decision schemes", *Psychological Review* 80, 97–125.

DeGroot, Morris (1974): "Reaching a Consensus", *Journal of the American Statistical Association* 69, 118–121.

Einhorn, Hillel J., Robin M. Hogarth, and Eric Klempner (1977): "Quality of Group Judgment", *Psychological Bulletin* 84, 158–172.

Elga, Adam. 2007. "Reflection and Disagreement". *Noûs* 41: 478–502.

Gigerenzer, Gerd, and Daniel G. Goldstein (1996): "Reasoning the fast and frugal way: Models of bounded rationality", *Psychological Review* 103, 650–669.

Hartmann, Stephan, and Jan Sprenger (2010): "The Weight of Competence Under a Realistic Loss Function", *The Logic Journal of the IGPL* 18, 346–352.

Henry, Rebecca A. (1995): "Improving Group Judgment Accuracy: Information Sharing and Determining the Best Member", *Organizational Behavior and Human Decision Processes* 62, 190–197.

Hill, Gayle W. (1982): "Group versus individual performance: Are N + 1 heads better than one?", *Psychological Bulletin* 91, 517–539.

Hinsz, Verlin B. (1999): "Group Decision Making with Responses of a Quantitative Nature: The Theory of Social Decision Schemes for Quantities", *Organizational Behavior and Human Decision Processes* 80, 28–49.

Hogarth, Robin M. (1978): "A Note on Aggregating Opinions", *Organizational Behavior and Human Performance* 21, 40–46.

Larrick, Richard P., Katherine A. Burson, and Jack B. Soll (2007): "Social comparison and overconfidence: When thinking you're better than average predicts overconfidence (and when it does not)", *Organizational Behavior and Human Decision Processes* 102, 76–94.

Laughlin, Patrick R., and Alan L. Ellis (1986): "Demonstrability and social combination processes on mathematical intellective tasks", *Journal of Experimental Social Psychology* 22, 177–189.

Lehrer, Keith, and Carl Wagner (1981): *Rational Consensus in Science and Society*. Reidel: Dordrecht.

Libby, Robert, Ken T. Trotman, and Ian Zimmer (1987): "Member variation, recognition of expertise and group performance", *Journal of Applied Psychology* 72, 81–87.

Lindley, Dennis V. (1983): "Reconcilation of probability distributions", *Operations Research* 31, 866–880.

Littlepage, Glenn E., Greg W. Schmidt, Eric W. Whisler, and Alan G. Frost (1995): "An input-process-output analysis of influence and performance in problem-solving groups", *Journal of Personality and Social Psychology* 69, 877–889.

List, Christian (2012): "The theory of judgment aggregation: an introductory review", *Synthese* 187, 179–207.

Martini, Carlo, Jan Sprenger, and Mark Colyvan (2013): "Consensus Through Mutual Respect: Scope and Limits", forthcoming in *Erkenntnis*.

Nadeau, Richard, Cloutier, Edouard, and Guay, J.-H. (1993): "New Evidence About the Existence of a Bandwagon Effect in the Opinion Formation Process", *International Political Science Review* 14, 203–213.

Page, Scott E. (2007): *The Difference*. Princeton: Princeton University Press.

Soll, Jack B., and Richard P. Larrick (2009): "Strategies for Revising Judgment: How (and How Well) People Use Others' Opinions", *Journal of Experimental Psychology* 35, 780–805.

Surowiecki, James (2004): *The Wisdom of the Crowds*. Harpswell: Anchor.

Thomas, Edwin J., and Clifton F. Fink (1961): "Models of group problem solving", *The Journal of Abnormal and Social Psychology* 63, 53–63.

Wilf, Herbert S. (1985): "Some Examples of Combinatorial Averaging", *American Mathematical Monthly* 92, 250–261.