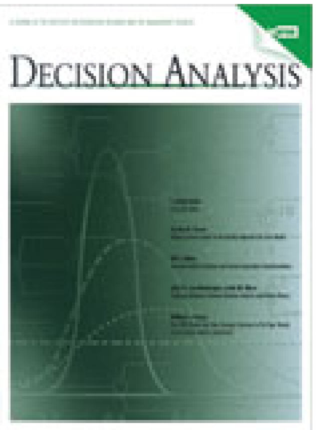


Decision Analysis

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Composition of Optimally Wise Crowds

Clintin P. Davis-Stober, David V. Budescu, Stephen B. Broomell, Jason Dana

DECISION ANALYSIS
Articles in Advance, pp. 1–14
ISSN 1545-8490 (print) | ISSN 1545-8504 (online)

http://dx.doi.org/10.1287/deca.2015.0315

The Composition of Optimally Wise Crowds

Clintin P. Davis-Stober
Department of Psychological Sciences, University of Missouri, Columbia, Missouri, 65211, stoberc@missouri.edu

David V. Budescu
Department of Psychology, Fordham University, Bronx, New York, 10458, budescu@fordham.edu

Stephen B. Broomell
Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213, broomell@cmu.edu

Jason Dana
School of Management, Yale University, New Haven, Connecticut, 06520, jason.dana@yale.edu

We investigate optimal group member configurations for producing a maximally accurate group forecast. Our approach accounts for group members that may be biased in their forecasts and/or have errors that correlate with the criterion values being forecast. We show that for large forecasting groups, the diversity of individual forecasts linearly trades off with forecaster accuracy when determining optimal group composition.

Keywords: type coherence; wisdom of the crowds; forecasting; prediction; optimal weighting; aggregation
History: Received on October 30, 2014. Accepted by Editor-in-Chief Rakesh K. Sarin on February 20, 2015, after 1 revision. Published online in *Articles in Advance*.

1. Introduction

There exists a large literature demonstrating that aggregating the forecasts/judgments of a group of individuals can yield surprisingly accurate results (for reviews, see Clemen 1989, Gigone and Hastie 1997, Wallsten et al. 1997). This phenomenon has been termed the "wisdom of the crowd" (Surowiecki 2004). Recent work on this topic has shown that smaller subsets of the crowd can be reliably more accurate than the crowd itself (Budescu and Chen 2015, Goldstein et al. 2014, Jose et al. 2014, Mannes et al. 2014). These "small crowds" benefit from the same principles that drive the general wisdom of the crowd effect, e.g., reduced variability in forecasts via averaging (e.g., Davis-Stober et al. 2014) but yield additional gains by restricting the crowd to the highest performing members (e.g., Budescu and Chen 2015, Goldstein et al. 2014, Mannes et al. 2014).

This body of work raises the question of whether we can determine the composition of an optimally accurate crowd. As an example, suppose an organization seeks to forecast the future performance of various publicly traded companies by employing M analysts. Should the organization simply select the analysts who have done well in the past and aggregate their responses? Or, drawing upon the literature on the benefits of forecaster diversity in groups (Hong and Page 2004), should the organization look to include individuals who yield systematically different forecasts from the other analysts, even if those individuals have a history of poorer individual accuracy?

We address these questions by developing a mathematical model that generalizes the one used by Lamberson and Page (2012; henceforth, LP). LP considered the composition of optimal forecasting groups who predict an unknown quantity, V. Their novel and somewhat counterintuitive result is that when all of M-many forecasters in a group are unbiased, the group's accuracy depends almost entirely on the diversity of individual forecasts and not their accuracy. In this paper, we extend LP's approach to examine optimal crowd composition under a richer definition of forecaster accuracy. First, we allow forecasters to be biased (i.e., the expectation of their forecasts is not necessarily V). Second, we allow their forecasts to covary with V (i.e., allow for correlated errors). Under these more general conditions, determining which individuals to include in a group becomes a more complex problem, and we show that the solution depends on three things: the particular source of individual (in)accuracy, diversity of individual forecasts, and group size.

There are many reasons to consider models in which individuals can be biased in their forecasts and have correlated errors with the variable being predicted. (1) Psychological motivations can lead people to bias their forecasts in the direction of their preferred outcomes. As examples, people tend to overestimate the likelihood that their favorite sports team will win a game (e.g., Simmons et al. 2011); their favorite candidate will win an election (e.g., Fischer and Budescu 1995); or, in corporate prediction markets, that a project on which they are working will be successful (Cowgill and Zitzewitz 2015). (2) Not all forecasters interpret information identically. Based on their background and training, some forecasters distort the impact of various sources of information. Good examples are found in studies of expert opinion elicitation in domains such as health (e.g., Wallsten et al. 1983) or climate change (e.g., Morgan et al. 2006), which find extreme disagreements among top experts. (3) Even without ideological biases, forecasters may exhibit path dependencies when processing information such that they fail to rationally update when evaluating evidence inconsistent with their initial impressions (e.g., Simon and Holyoak 2002). (4) Forecasters may have different subsets of relevant information for making an accurate forecast and the pattern of missing information is typically not random. Some information is more expensive, harder to interpret, or simply unavailable, and these predictable patterns of missing information can bias some forecasters. Any of these reasons, or combinations thereof, can introduce bias into forecasts. Further, it is not always possible to correct for biases. Forecast biases could be present but change dynamically depending upon either known or unknown variables (e.g., Massey et al. 2011, Sinclair et al. 2010). In such cases, one could bound individual biases but not eliminate them from the aggregated forecast.

Our model offers a general set of guidelines for forming optimal groups. The following four guidelines emerge as general principles for optimal crowd formation.

- Rule 1: Differences in forecast variance among individual group members matters less as the group becomes larger.
- Rule 2: Individual forecast biases are important no matter what the size of the group but become even more important in larger groups with larger biases leading to less accurate groups.
- Rule 3: Individual forecast covariance with the criterion being predicted becomes more important as the group becomes larger.
- Rule 4: Diversity of individual forecasts becomes more important as the group becomes larger, with more diversity leading to more accurate groups.

The core contribution of our work is to precisely state how the diversity and accuracy of individual forecasts determine optimally accurate groups. We show that for large forecasting groups, the diversity of individual forecasts linearly trades off with forecaster accuracy when determining optimal group composition.

1.1. Preliminary Assumptions and Definitions

As in LP, assume there are M-many forecasters and let s_i denote the realization of forecaster i. Let $\text{var}(s_i)$ denote forecaster i's variance. Note that when the criterion variable being forecast, defined as V, is a fixed, nonrandom quantity, the variance of forecaster i is his/her error variance as defined in LP; i.e., $\text{var}(s_i) = \text{var}(s_i - V) = \text{var}(\epsilon_i)$, where $\text{var}(\epsilon_i)$ is the forecast variance in the notation of LP. Let Σ denote the covariance matrix of the M-many forecasters. Let w_i, $i \in \{1, 2, \ldots, M\}$ denote a set of real-valued, nonnegative weights such that $\sum_{i=1}^{M} w_i = 1$. We consider weighted linear combinations of forecasts, $G(s) = \sum_{i=1}^{M} w_i s_i$. The assumption of nonnegative weights is tantamount to assuming that each forecaster being aggregated has positive predictive accuracy at the forecasting task; i.e., their forecasts are not negatively correlated with the criterion variable.

Our generalization of LP mirrors the modeling framework in Davis-Stober et al. (2014). First, we treat V as a real-valued random variable with finite expectation and variance, denoted $E[V]$ and $\text{var}(V)$, respectively. Since V is now allowed to be random, the covariance between V and the individual forecasters is well defined. Let $\text{cov}(s_i, V)$ be defined as the covariance between the ith forecaster and the criterion variable. From the perspective of forecaster error, the $\text{cov}(s_i, V)$ term allows us to consider the case where a forecaster's errors are correlated with V. Whether V should be considered a random variable or a fixed quantity will, of course, depend upon the substantive question at hand. We note that for all of our results, the case of nonrandom V can be obtained by simply setting $\text{var}(V)$ and all $\text{cov}(s_i, V)$ terms equal to zero. Second, we allow the forecasters to be biased in their predictions. Define $\delta_i = E[s_i - V]$ as forecaster i's *bias*. Without loss of generality, we let $E[V] = 0$; i.e., we assume that the criterion variable V is centered at 0 and hence $\delta_i = E[s_i]$.

As in LP, we evaluate accuracy via *mean squared error*, defined as the expected squared difference between the group aggregate and the criterion. In LP, individual forecaster accuracy is modeled simply as $\text{var}(s_i)$. Under our formulation, forecaster accuracy has a richer definition that includes variance, $\text{var}(s_i)$, bias, δ_i, and forecaster covariance with the criterion variable, $\text{cov}(s_i, V)$. These three terms trade off with each other in interesting ways when considering individual forecaster accuracy. For example, via the well-known bias-variance trade-off, a biased forecaster may be more accurate on average than an unbiased forecaster by having less variance (or higher covariance with V) in his or her forecasts. We now re-derive each result from LP, using the same theorem and corollary numberings. Each result contains the corresponding result from LP as a special case. All proofs are in Appendix A.

Theorem 1. *The expected squared error of $G(s)$ when G gives the simple average of the signals $s_1, \ldots, s_M$ is*

$$E[(G(s)-V)^2] = \overline{\text{bias}}(s)^2 + \frac{1}{M}\overline{\text{var}}(s) + \left(1-\frac{1}{M}\right)\overline{\text{cov}}(s) - 2\overline{\text{cov}}(s, V) + \text{var}(V),$$

where $\overline{\text{bias}}(s) = (1/M)\sum_{i=1}^{M}\delta_i$ *(average forecaster bias),* $\overline{\text{var}}(s) = (1/M)\sum_{i=1}^{M}\text{var}(s_i)$ *(average forecaster variance),* $\overline{\text{cov}}(s) = (1/(M(M-1)))\sum_{i=1}^{M}\sum_{i\neq j}\text{cov}(s_i, s_j)$ *(average covariance among forecasters), and* $\overline{\text{cov}}(s, V) = (1/M)\sum_{i=1}^{M}\text{cov}(s_i, V)$ *(average covariance between forecasters and criterion, which we refer to as "forecasters' predictive validity").*

Theorem 1 reduces to LP's Theorem 1 when $\text{cov}(s_i, V)$ and $\text{var}(V)$ are set to zero. The larger the covariances between the criterion variable and the individual forecasts, the more accurate the group's aggregate. Also, large positive covariances between the criterion variable and individual forecasters, $\text{cov}(s_i, V)$, can compensate for biased predictions, large inter-forecaster covariances, and/or large forecaster variances. The key insight to be gained from Theorem 1 is that all aspects of crowd wisdom (e.g., forecaster bias, covariance) linearly trade off with one another when assessing accuracy. In other words, high forecaster accuracy (low bias, low forecaster variance, high covariance with criterion variable) can compensate for large amounts of positive covariation among forecasters, i.e., low diversity.

2. Optimal Group Weighting and Composition

Although Theorem 1 was derived under the assumption that the group aggregate is a simple average, it is instructive to consider the *optimal* aggregation weights. The following theorem provides this solution.

Theorem 2. *The optimal weighted average of a collection of signals $s = (s_1, \ldots, s_M)$ with covariance matrix Σ is the solution to the following system of linear equalities:*

$$\begin{pmatrix} \mathbf{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}' & \mathbf{u} \\ \mathbf{u}' & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{s,V} \\ 1 \end{pmatrix},$$

where $\mathbf{u}$ is a $M \times 1$ vector of ones, $\boldsymbol{\delta}$ is the $M \times 1$ vector of signal biases, $\boldsymbol{\sigma}_{s,V}$ is the $M \times 1$ vector of covariances between each signal and the criterion variable V, $\boldsymbol{w}$ is the $M \times 1$ vector of weights constrained to sum to 1, and λ is a real-valued unknown variable. By solving the above system for $\mathbf{w}$, we obtain the optimal set of weights $\mathbf{w}^*$.

If the matrix $\mathbf{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}'$ is positive definite, then the optimal set of weights always exists and is unique. The optimal set of weights will necessarily be a function of both "accuracy" terms ($\text{cov}(s_i)$, $\text{var}(s_i)$, δ_i) as well as "diversity terms" (the off-diagonal elements of Σ).

Example 1. As a simple numerical example, consider a small group composed of six individuals. Let the covariance matrix of the members' forecasts be as follows:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 & 0.15 & 0.15 & 0 & 0 \\ 0.5 & 1 & 0.15 & 0.15 & 0 & 0 \\ 0.15 & 0.15 & 1 & 0.25 & 0 & 0 \\ 0.15 & 0.15 & 0.25 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.10 \\ 0 & 0 & 0 & 0 & 0.10 & 1 \end{pmatrix},$$

with a corresponding bias vector equal to

$$\boldsymbol{\delta}' = \begin{pmatrix} 0 & 0 & 0.2 & 0.2 & 0.5 & 0.5 \end{pmatrix},$$

and covariance vector with the criterion variable, V, equal to

$$\boldsymbol{\sigma}'_{s,V} = \begin{pmatrix} 0.3 & 0.3 & 0.35 & 0.35 & 0.25 & 0.25 \end{pmatrix}.$$

This numerical example illustrates the trade-off in accuracy and diversity when determining optimal group weighting. For this example, we assume all forecast variances to be equal to one. The group members corresponding to the first two rows of $\boldsymbol{\Sigma}$ produce highly correlated forecasts, and these forecasts positively correlate with the next two members, thus producing the least amount of diversity in the group. The next two members are similar but have predictions that inter-correlate less so than the first two. Finally, the last two members have the smallest inter-correlation and produce forecasts that are uncorrelated with the other two members. These final two members produce the largest amount of diversity in the group. Examining the bias vector, we see that the two members with the lowest diversity are unbiased in their predictions. Conversely, the two group members that produce the most diverse forecasts in the group are also the most biased. Finally, the vector of covariances with the criterion are similar in magnitude across the six members.

Intuitively, it seems sensible to heavily weight the two most accurate (unbiased) members and place less positive weight, if any, on the most biased. Yet the two most biased group members also produce the most diverse forecasts because their forecasts are uncorrelated with the other members. Theorem 2 provides a method for determining the best way to weight and combine the forecasters taking into account these trade-offs. Applying Theorem 2, we carry out the matrix algebra and optimal an optimal weight vector of

$$\mathbf{w}^* = \begin{pmatrix} 0.16 & 0.16 & 0.21 & 0.21 & 0.13 & 0.13 \end{pmatrix}.$$

This result is interesting because the two unbiased judges did not receive the largest weights. The high inter-correlation (low diversity) of those two judges resulted in a smaller weighting. Also, the most biased group members did receive the lowest weights, but these weights are still reasonably large, reflecting the value of the diverse forecasts. Finally, the remaining two forecasters received the largest weights, reflecting their more balanced forecast behavior (moderate diversity, moderate bias). Next, as in LP, it is instructive to consider the simple case of two forecasters.

Example 2 (Two Forecasters). Suppose that two forecasters, a and b, have a forecast covariance matrix equal to

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{var}(a) & \mathrm{cov}(a,b) \\ \mathrm{cov}(a,b) & \mathrm{var}(b) \end{pmatrix};$$

a bias vector equal to

$$\boldsymbol{\delta} = \begin{pmatrix} \delta_a \\ \delta_b \end{pmatrix};$$

and covariance vector with the criterion variable, V, equal to

$$\boldsymbol{\sigma}_{\mathrm{s,V}} = \begin{pmatrix} \mathrm{cov}(a,V) \\ \mathrm{cov}(b,V) \end{pmatrix}.$$

Then the optimal weights, given by Theorem 2, are

$$\frac{\delta_b^2 - \delta_a\delta_b + \mathrm{cov}(a,V) - \mathrm{cov}(b,V) + \mathrm{var}(b) - \mathrm{cov}(a,b)}{(\delta_a - \delta_b)^2 + \mathrm{var}(a) + \mathrm{var}(b) - 2\mathrm{cov}(a,b)},$$

for forecaster a and

$$\frac{\delta_a^2 - \delta_a\delta_b + \mathrm{cov}(b,V) - \mathrm{cov}(a,V) + \mathrm{var}(a) - \mathrm{cov}(a,b)}{(\delta_a - \delta_b)^2 + \mathrm{var}(a) + \mathrm{var}(b) - 2\mathrm{cov}(a,b)},$$

for forecaster b. We can further simplify the denominator for both weights, yielding

$$\frac{\delta_b^2 - \delta_a\delta_b + \mathrm{cov}(a,V) - \mathrm{cov}(b,V) + \mathrm{var}(b) - \mathrm{cov}(a,b)}{\mathrm{MSE}(a-b)},$$

for a and

$$\frac{\delta_a^2 - \delta_a\delta_b + \mathrm{cov}(b,V) - \mathrm{cov}(a,V) + \mathrm{var}(a) - \mathrm{cov}(a,b)}{\mathrm{MSE}(a-b)},$$

for b, where $\text{MSE}(a-b)$ is the mean squared error of the difference between forecaster a and b. The optimal weighting is a function of forecaster biases and their covariances with V, in addition to elements of Σ. Also, whenever two of the three features (bias, variance, covariance with V) are set equal for both a and b, the weights become a simple linear function of the third. For example, if one assumes equal bias and equal variances, the two weights are linear functions of the covariances with V.

This simple example provides insight into the nature of the optimal weights. If both forecasters have equal bias, then bias plays no role in determining optimal group weighting, as one would expect. Also, the difference between the covariances of the forecasters, $\text{cov}(a, V) - \text{cov}(b, V)$, plays an equal role in determining the optimal weighting as the difference between the variance of the forecaster and the covariance in predictions between both forecasters, e.g., $\text{var}(b) - \text{cov}(a, b)$. Clearly, a forecaster will be more heavily weighted than the other if he/she has less bias than the other forecaster, has a higher covariance with the criterion, and has predictions that do not largely covary with the other forecaster.

Next, we will consider optimal group composition under the "type coherence" structure. LP define two forecaster types, a and b, and all forecasts made by type a (respectively, type b) have equal variances, $\text{var}(a)$ (respectively, $\text{var}(b)$). Let $\text{cov}(a)$ (respectively, $\text{cov}(b)$) denote the covariance in the forecasts made by type a (respectively, type b) forecasters, and assume that this is identical for any pair of type a (respectively, type b) forecasters. Let $\text{cov}(a, b)$ denote the covariance in the forecasts of any two forecasters, one that is type a and the other type b. As in LP, the two types, a and b, satisfy *type coherence* if $\text{TC}(a, b) > 0$, where $\text{TC}(a, b) = \text{cov}(a) + \text{cov}(b) - 2\text{cov}(a, b)$. Our generalization will assume that all type a judges have equal bias, δ_a, and all type b judges have equal bias, δ_b. Likewise, we assume that all type a judges have identical covariance with V, $\text{cov}(a, V)$, and all type b judges have identical covariance with V, $\text{cov}(b, V)$. From our Theorem 1, the expected squared error of a simple average of a group with A-many type a forecasters and B-many type b forecasters is given by the following expression where $M = A + B$:

$$
\begin{aligned}
E[(G(s)-V)^2] &= \frac{1}{M^2}\cdot(A(\delta_a^2+\text{var}(a))+B(\delta_b^2+\text{var}(b)) \\
&\qquad +A(A-1)(\delta_a^2+\text{cov}(a))+B(B-1)(\delta_b^2+\text{cov}(b)) \\
&\qquad +2AB(\delta_a\delta_b+\text{cov}(a,b))) \\
&\quad -\frac{2(A\text{cov}(a,V)+B\text{cov}(b,V))}{M}+\text{var}(V).
\end{aligned}
$$

Theorem 3. *In a group of size M the optimal fraction of type a forecasters is approximated by*

$$
\begin{aligned}
&\frac{[\text{var}(b)-\text{var}(a)]-[\text{cov}(b)-\text{cov}(a)]}{2M(\text{TC}(a,b)+(\delta_a-\delta_b)^2)} \\
&\quad +([\text{cov}(b)-\text{cov}(a,b)]+[\delta_b^2-\delta_a\delta_b] \\
&\quad +[\text{cov}(a,V)-\text{cov}(b,V)]) \\
&\quad \cdot(\text{TC}(a,b)+(\delta_a-\delta_b)^2)^{-1},
\end{aligned} \tag{1}
$$

where the approximation error is less than $1/M$. The optimal fraction can equivalently be restated in terms of correlations, yielding

$$
\begin{aligned}
&\frac{[\text{var}(b)-\text{var}(a)]-[\text{var}(b)\text{corr}(b)-\text{var}(a)\text{corr}(a)]}{2M(\text{TC}(a,b)+(\delta_a-\delta_b)^2)} \\
&\quad +\frac{[\text{var}(b)\text{corr}(b)-\sqrt{\text{var}(a)}\sqrt{\text{var}(b)}\text{corr}(a,b)]}{\text{TC}(a,b)+(\delta_a-\delta_b)^2} \\
&\quad +([\delta_b^2-\delta_a\delta_b]+[\sqrt{\text{var}(a)}\sqrt{\text{var}(V)}\text{corr}(a,V) \\
&\qquad -\sqrt{\text{var}(b)}\sqrt{\text{var}(V)}\text{corr}(b,V)]) \\
&\quad \cdot(\text{TC}(a,b)+(\delta_a-\delta_b)^2)^{-1},
\end{aligned}
$$

where $\text{corr}(\cdot)$ *denotes correlation.*

Example 3. As a simple example, consider a population of type a and type b forecasters under the following assumptions. Let $\text{var}(a) = 1, \text{var}(b) = 2$, $\text{cov}(a) = 0.6, \text{cov}(b) = 0.3, \text{cov}(a, b) = 0.10, \delta_a = 0.1$, $\delta_b = 0.5, \text{cov}(a, V) = 0.35$, $\text{cov}(b, V) = 0.15$, and $\text{var}(V) = 1.5$. The respective correlations are $\text{corr}(a) = 0.6, \text{corr}(b) = 0.15, \text{corr}(a, b) = 0.07, \text{corr}(a, V) = 0.29$, and $\text{corr}(b, V) = 0.087$. Here, the type a forecasters are more accurate than type b because they have smaller biases, lower forecast variances, and higher covariances with the criterion variable. Conversely, type b forecasters are more diverse because they have lower inter-type covariances than type a. Next, we calculate the optimal fraction of type a forecasters as a function of group size M, using Theorem 3. For $M = 10$, the optimal fraction of type a forecasters is 0.77, reflecting the comparatively larger accuracy advantage of type a. For $M = 20$, the optimal fraction is 0.73. For $M = 100$, the optimal fraction is 0.70. As can be seen in Equation (1), as M becomes larger, the impact of differences in variances becomes smaller, giving a slight advantage to type b forecasters.

The right-hand term of (1), which is not a function of group size, M, contains several terms relating to accuracy—forecaster biases and forecaster covariances with the criterion variable, V. In other words, in contrast to LP, accuracy remains important in determining optimal group composition in our Theorem 3, regardless of group size. As we see in the next result, forecaster variance becomes negligible in large groups, as in LP, but forecaster bias and forecaster covariance with the criterion variable do not. This necessarily impacts all remaining results regarding small and large group optimal composition.

Theorem 4. *As the group size M approaches infinity, the optimal fraction of type a forecasters approaches*

$$([\mathrm{cov}(b)-\mathrm{cov}(a,b)]+[\delta_b^2-\delta_a\delta_b] + [\mathrm{cov}(a,V)-\mathrm{cov}(b,V)])\cdot(\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2)^{-1}. \quad (2)$$

This optimal fraction can equivalently be restated in terms of correlations, yielding

$$\frac{[\mathrm{var}(b)\mathrm{corr}(b)-\sqrt{\mathrm{var}(a)}\sqrt{\mathrm{var}(b)}\mathrm{corr}(a,b)]}{\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2} + ([\delta_b^2-\delta_a\delta_b]+[\sqrt{\mathrm{var}(a)}\sqrt{\mathrm{var}(V)}\mathrm{corr}(a,V) - \sqrt{\mathrm{var}(b)}\sqrt{\mathrm{var}(V)}\mathrm{corr}(b,V)]) \cdot(\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2)^{-1}.$$

Example 4. Returning to the set of assumptions used in Example 3, we can use Theorem 4 to determine the optimal fraction of type a forecasters for arbitrarily large groups. Applying (2), we obtain 0.689 as the optimal fraction of type a forecasters. Although diversity continues to be important for such large groups, the accuracy advantage of type a continues to play a significant role.

The limiting fraction in (2) depends on the covariances among forecasters as well as forecaster bias and forecaster covariances with the criterion. As in LP, the extent to which type a forecasters make systematically different predictions than type b forecasters make, as captured by the term $\mathrm{cov}(b)-\mathrm{cov}(a,b)$, continues to play an important role in determining the optimal configuration of type a forecasters; i.e., the "diversity" of forecasts is still important. However, the role of these covariance terms is offset by the difference in bias, $\delta_b(\delta_b-\delta_a)$, as well as the difference in covariances between the criterion variable and type a and b forecasters. Note that what is important with regard to bias is not whether type a and b forecasters are unbiased, per se, only the extent to which they are biased *differently*.

Equation (2) provides a clear interpretation of optimal crowd composition. Let $\delta_b(\delta_b-\delta_a)$ denote the *bias advantage* of type a. Let $\mathrm{cov}(a,V)-\mathrm{cov}(b,V)$ denote the *responsive advantage* of type a.[1] Let $\mathrm{cov}(b)-\mathrm{cov}(a,b)$ denote the *diversity advantage* of type a. We can restate Equation (2) as follows. As the group size M approaches infinity the optimal fraction of type a forecasters approaches

$$(\text{Diversity advantage}+\text{Bias advantage} + \text{Responsive advantage}) \cdot(\text{Type coherence}+\text{Differences in Bias-Squared})^{-1}.$$

It is perhaps surprising that all three difference terms are linearly related. In other words, differences between the biases of the two types can offset any difference between forecaster covariances. From this perspective, accuracy directly, and linearly, trades off with diversity for very large groups.

In Figure 1, we show the impact of group size, M, and the difference in bias between forecaster types, $\delta_b-\delta_a$, on the optimal fraction of type a forecasters. Forecaster covariances with the criterion are removed from this analysis by setting $\mathrm{cov}(a,V)-\mathrm{cov}(b,V)=0$. We examine three scenarios where (1) there is a negative bias for the type b forecasters, (2) there is no bias for the type b forecasters, and (3) there is a positive bias for the type b forecasters. The optimal fraction of type a forecasters is computed from Theorem 3. The results from LP's Figure 1 are replicated in panel (b) where $\delta_b-\delta_a=0$. For panel (a), when type b is biased

[1] If $\mathrm{var}(a)=\mathrm{var}(b)$, the responsive advantage can be interpreted as a difference in correlations multiplied by $\mathrm{var}(a)$.

at $\delta_b = -5$, there is a large span of δ_a values such that the optimal group should be entirely composed of type a forecasters. Finally, all three graphs in Figure 1 demonstrate that the impact of M on the optimal fraction of type a forecasters is maximal when both type a and b forecasters are unbiased. The remaining curves have less slope for all other values for $\delta_b - \delta_a$, as a function of M, and in some cases M is irrelevant.

We can now state LP's Corollary 1 more generally.

Corollary 1. *If both forecasting types have identical biases and covariances with the criterion, then for sufficiently small groups, the lowest variance type should be in the majority. Equivalently, if both forecasting types have identical biases and covariances with the criterion, the forecaster type with the highest correlation with the criterion variable should be in the majority.*

LP's Corollary 1 holds only under the restrictive case when both types have identical biases as well as identical covariances with the criterion. Otherwise, such accuracy versus diversity recommendations will necessarily depend upon group size, magnitude of difference in the biases between type a and type b forecasters, forecaster variances and covariances, and so on. We now solve for the precise nature of this trade-off, for both small and large groups, in the form of optimal group configuration.[2]

Theorem 5. *Suppose that the forecasts of type a and type b forecasters are independent of one another, i.e.,* $\mathrm{cov}(a, b) = 0$, *and that* $\mathrm{cov}(b) < \mathrm{cov}(a)$. *Then a group of size M should contain a majority of type b forecasters if*

$$\frac{[\delta_a^2 - \delta_b^2] + 2[\mathrm{cov}(b, V) - \mathrm{cov}(a, V)] + [\mathrm{cov}(a) - \mathrm{cov}(b)]}{[\mathrm{var}(b) - \mathrm{var}(a)] + [\mathrm{cov}(a) - \mathrm{cov}(b)]} < \frac{1}{M}.$$

Under LP's assumptions, i.e., unbiased judges and a nonrandom criterion variable, V, the above theorem simplifies to[3]

$$\frac{\mathrm{cov}(a) - \mathrm{cov}(b)}{\mathrm{var}(b) - \mathrm{var}(a)} < \frac{1}{M-1}.$$

[2] The main result of Theorem 5 can be written in terms of correlations by substituting the appropriate terms (as in Theorems 3 and 4).

[3] This differs slightly from LP in that they report $1/(M+1)$ on the right-hand side of this inequality.

Our generalized Theorem 5 provides a different substantive conclusion than LP concerning the optimal trade-offs between accuracy and diversity. LP state (Lamberson and Page 2012, p. 808) that "Thus, the threshold ratio of the difference between within-type covariances and the type variances increases linearly in group size; if the group size becomes twice as large, the ratio of differences in covariances to differences in accuracies needs only be half as large." Theorem 5 shows that this conclusion does not hold in the general case. Next, we more fully explore the precise nature of these trade-offs as a function of optimal group composition.

3. Relative Importance of Type Covariance, Validity, and Bias

By applying Theorem 4, we analyze the relative importance of the differences between forecaster types with regard to bias, covariance with the criterion, and within- and between-type covariance. As the size of the group approaches infinity, the optimal fraction of type a forecasters approaches

$$F_a = ([\mathrm{cov}(b) - \mathrm{cov}(a, b)] + \delta_b[\delta_b - \delta_a] + [\mathrm{cov}(a, V) - \mathrm{cov}(b, V)]) \cdot (\mathrm{TC(a, b)} + (\delta_b - \delta_a)^2)^{-1}.$$

We can re-express this quantity in terms of changes in forecaster covariances, bias, and criterion covariance. Let $\Delta_{C_b} = [\mathrm{cov}(b) - \mathrm{cov}(a, b)]$ and $\Delta_{C_a} = [\mathrm{cov}(a) - \mathrm{cov}(a, b)]$ be the difference of within- and between-type covariances. Let $\Delta_B = [\delta_b - \delta_a]$ be the difference in the bias for type a and b forecasters. Let $\Delta_V = [\mathrm{cov}(a, V) - \mathrm{cov}(b, V)]$ be the difference in the criterion covariance for type a and b forecasters. We can re-express the optimal fraction of type a forecasters as

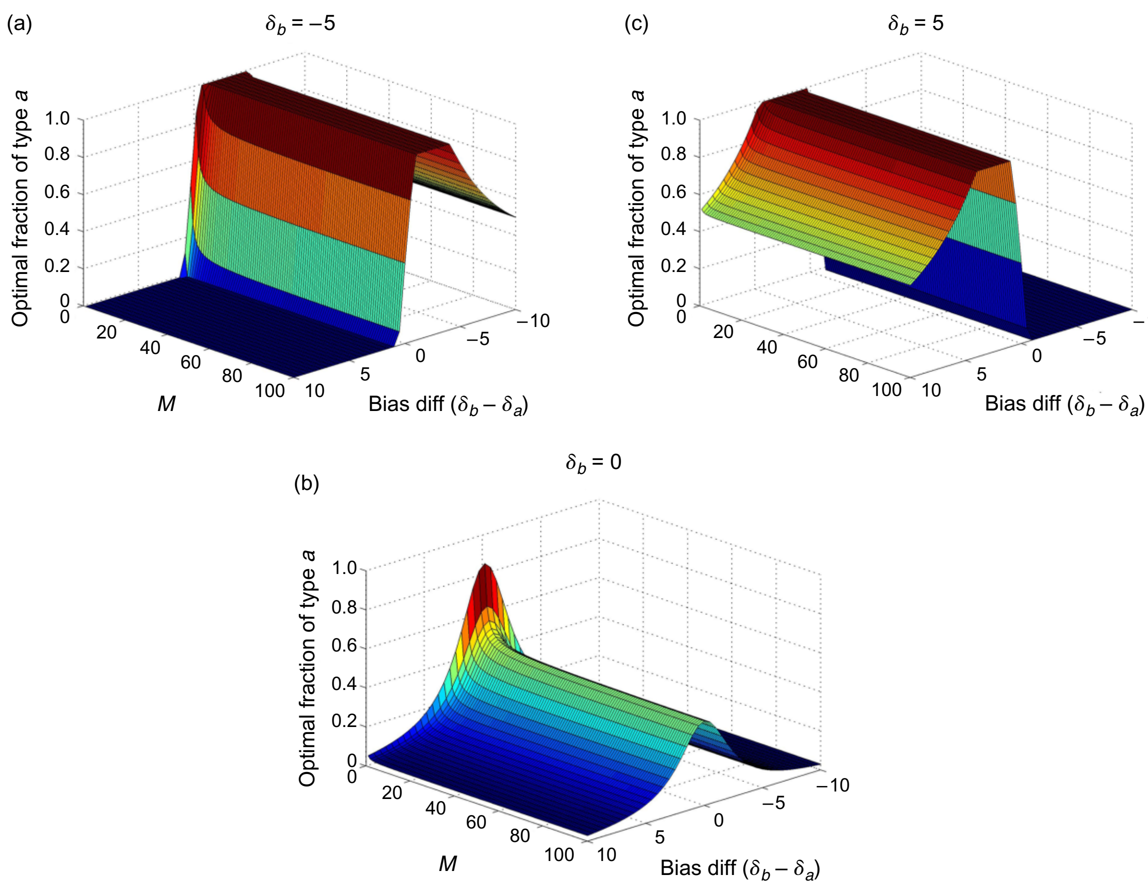
$$F_a = \frac{[\Delta_{C_b}] + \delta_b[\Delta_B] + [\Delta_V]}{\Delta_{C_a} + \Delta_{C_b} + (\Delta_B)^2}.$$

To see more clearly how these differences trade off with one another in determining optimal group size, we take the partial derivative of F_a with respect to each of the Δ variables described above. Please see Appendix B for the mathematical details.

First, consider the relative importance of differences in bias versus that of criterion covariance. The difference in bias, Δ_B, has more impact than difference in criterion covariance, Δ_V, when

$$\frac{\partial F_a}{\partial \Delta_B} > \frac{\partial F_a}{\partial \Delta_V}.$$

Figure 1 (Color online) The Optimal Fraction of Type a Forecasters Plotted Against the Size of the Group, M, and the Difference in Bias Between the Types, $(\delta_b - \delta_a)$



Notes. These results occur when assuming $\text{var}(a) = 5$, $\text{var}(b) = 10$, $\text{cov}(a) = 2$, $\text{cov}(b) = 1$, and $\text{cov}(a, b) = -2$. Panels (a)–(c) are plotted assuming $\delta_b = -5, 0, 5$, respectively.

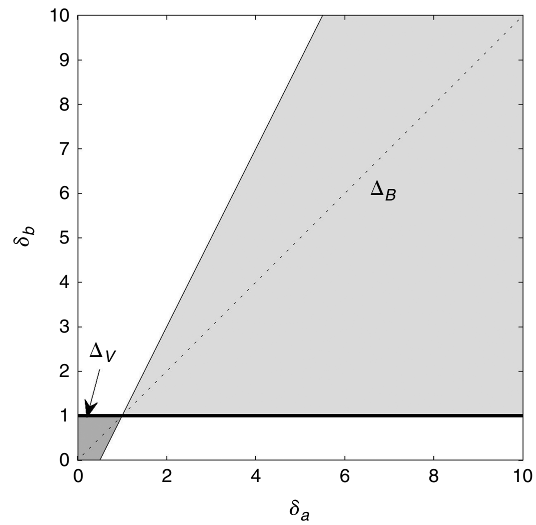
Solving for F_a, which is bounded on the interval $[0, 1]$, we get the following results.

$$F_a < \frac{\delta_b - 1}{2(\delta_b - \delta_a)} \quad \text{when } \delta_b > \delta_a. \tag{3}$$

$$F_a > \frac{\delta_b - 1}{2(\delta_b - \delta_a)} \quad \text{when } \delta_b < \delta_a. \tag{4}$$

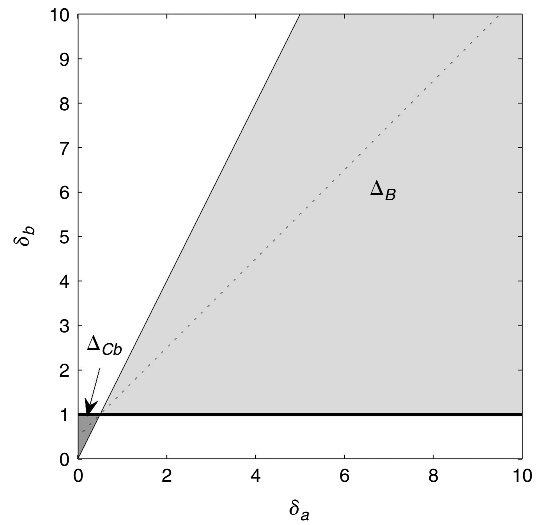
The relative importance of the difference of bias versus the difference of criterion covariance is a function of F_a, δ_a, and δ_b. We can plot the value of the right-hand side (RHS) of Inequalities 3 and 4 in the Cartesian plane defined by δ_a and δ_b. Since the optimal fraction, F_a, is restricted to the unit interval $[0, 1]$, any values of the RHS of Inequalities 3 and 4 falling inside this unit interval are labeled indeterminate because the most important variable depends on the value of the optimal fraction, F_a. Values of the RHS of Inequalities 3 and 4 that fall outside the unit interval completely determine the more influential variable. Therefore, in Figure 2 we can divide the Cartesian space defined by δ_a and δ_b into three regions based on which variable has more impact on the optimal fraction: (a) bias difference is more important, (b) criterion covariance difference is more important, and (c) difference is indeterminate. The dotted line identifies cases where bias for type a and b forecasters is equal. The thin solid line indicates when the RHS is equal to one. The thick solid line indicates where the ratio on the RHS is zero. The light shaded area is the region where Δ_B has more impact, and the dark shaded area indicates the region where Δ_V has more

Figure 2 The Region of the Cartesian Space Defined by δ_a and δ_b Where the Difference in Bias Between Forecaster Types Is More Influential Than the Difference in Criterion Covariance Between Forecaster Types



Note. The light shaded region favors bias, and the dark shaded region favors criterion covariance.

Figure 3 The Region of the Cartesian Space Defined by δ_a and δ_b Where the Difference in Bias Between Forecaster Types Is More Influential Than the Difference in Covariance Between Forecaster Types



Note. The light shaded region favors bias, and the dark shaded region favors covariance.

impact. The white regions are indeterminate, however, in the upper left-hand corner of the plot, the ratio on the RHS equal to 0.5 or more, and in the lower left-hand corner of the plot, the ratio is only slightly greater than zero.

This analysis reveals that the impact of the difference in bias (Δ_B) is relative to the overall amount of bias (δ_b). If the magnitude of bias (say, for type *b* forecasters) is large, the difference in bias between type *a* and *b* forecasters (Δ_B) needs to be large in order have more influence on the optimal fraction than the difference in criterion covariance between type *a* and *b* forecasters (Δ_V). If the magnitude of bias is smaller, smaller differences in bias are more meaningful. In general, the magnitude of bias can be interpreted relative to the variances of the forecasts and the criterion. For example, if all the variances were equal to one, the individual bias terms could be interpreted in units of standard deviations of the criterion variable (assuming that the criterion variable is centered at zero).

Similarly, the difference in bias, Δ_B, has more impact than the difference, Δ_{C_b}, when

$$\frac{\partial F_a}{\partial \Delta_B} > \frac{\partial F_a}{\partial \Delta_{C_b}}.$$

Solving again for F_a we get the following results.

$$F_a < \frac{\delta_b - 1}{2(\delta_b - \delta_a) - 1} \quad \text{when } 2(\delta_b - \delta_a) - 1 > 0. \tag{5}$$

$$F_a > \frac{\delta_b - 1}{2(\delta_b - \delta_a) - 1} \quad \text{when } 2(\delta_b - \delta_a) - 1 < 0. \tag{6}$$

The relative importance of the bias difference versus Δ_{C_b} is a similar function of F_a, δ_a, and δ_b as in the previous analysis. Figure 3 is similar to Figure 2 but shows some important differences. The area of the region where the difference in within- and between-type covariances is more important than the bias difference is small, and we can see that Δ_{C_b} is more influential than Δ_B at the point where $\delta_a = \delta_b = 0$, reflecting the results of LP with no bias.

Finally, the difference in criterion covariance, Δ_V, has more impact than Δ_{C_b} when

$$\frac{\partial F_a}{\partial \Delta_V} > \frac{\partial F_a}{\partial \Delta_{C_b}}.$$

Solving again for F_a we get the following results.

$$F_a > 0.$$

Here we can see that the difference in criterion covariance is always more important than Δ_{C_b} unless the optimal fraction for type *a* forecasters is equal to zero.

In general, differences in bias and criterion covariance are extremely important in determining the optimal group composition. The importance of differences in bias increases with the amount of bias for each forecaster type, showing a much larger region where differences in bias are the most influential determinant of the optimal fraction of forecasters. Next, as in LP, we consider the conditions under which the optimal group configuration contains both type a and b forecasters.[4]

Theorem 6. *The optimal group of size M contains both type a and type b forecasters if and only if*

$$\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2-M([\mathrm{cov}(b)-\mathrm{cov}(a,b)] + [\delta_b^2-\delta_a\delta_b]+[\mathrm{cov}(a,V)-\mathrm{cov}(b,V)]) \le F(a,b), \quad (7)$$

and

$$F(a,b) \le (M-1)(\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2) - M([\mathrm{cov}(b)-\mathrm{cov}(a,b)]+[\delta_b^2-\delta_a\delta_b] + [\mathrm{cov}(a,V)-\mathrm{cov}(b,V)]), \quad (8)$$

where $F(a,b)=([\mathrm{var}(b)-\mathrm{var}(a)]-[\mathrm{cov}(b)-\mathrm{cov}(a)])/2.$

The left-hand side of Inequality (7) can be described as the difference between two terms. The first term is the sum of type coherence and squared difference in bias between the two types. The second term is the sample size multiplied by the sum of the relative advantages of type a, i.e., the sum of responsive advantage, bias advantage, and diversity advantage. Inequality (8) has a similar structure, with the right-hand side being the difference between the same two terms with the exception that now the first term is multiplied by $M-1$. The interpretation of these two inequalities is that for the optimal group to have both type a and type b forecasters, the sum of the relative advantages of type a, multiplied by sample size, M, should not completely outweigh the benefit of the additional diversity of type b forecasters.

Again, type biases and covariances continue to trade off with forecaster variance and covariances. As expected, an optimal group should contain both forecaster types as long as the group types provide some diversity in prediction (forecasts between the types do not perfectly co-vary) without one group type incurring a corresponding large amount of bias and/or weak covariance with the criterion variable, V.

LP's conclusions regarding strong type coherence become more nuanced under our generalization. *Strong type coherence* is defined as the case where $\mathrm{cov}(a) > \mathrm{cov}(a,b)$ and $\mathrm{cov}(b) > \mathrm{cov}(a,b)$. Under the assumptions of unbiased forecasters and nonrandom V, the conditions laid out in Theorem 6 always hold under the case of strong type coherence. Under our more general formulation, strong type coherence is not sufficient to guarantee the inclusion of both forecaster types in the optimal group configuration; whether both types should be included in the optimal group composition now also depends upon type biases and covariances with the criterion variable.

4. Discussion

4.1. Modeling Signals versus Signal Errors

As in LP, we measured group accuracy via the mean squared error between linear combinations of signals, s_i, and the criterion V. In our subsequent derivations, we considered forecaster variances and covariances at the level of the signals, s_i, as opposed to the LP formulation, which operated at the level of the signal errors, $s_i - V$. Under the assumption of nonrandom V and unbiased judges, all variance and covariance definitions are identical, e.g., $\mathrm{var}(s_i)=\mathrm{var}(s_i-V)$; hence our results contain LP's as a special case. Similarly, our results could also be rewritten solely in terms of errors as opposed to signals.

LP's original results (Theorems 2–6) can easily be generalized to the case of biased individual forecasters with a nonrandom V. If we allow biased forecasters, then the variance of the signal errors is equal to $\mathrm{var}(e_i)=\delta_i^2+\mathrm{var}(s_i)$, and the covariance between individual errors is equal to $\mathrm{cov}(e_i,e_j)=\delta_i\delta_j+\mathrm{cov}(s_i,s_j)$. In this way, it is straightforward to show that LP's Theorems 2–6 do, in fact, generalize to the case of biased forecasters under nonrandom V. However, as we have just shown, these error variances and covariances necessarily contain bias terms. This is important because the covariance among error terms can no

[4] The main result of Theorem 6 can be written in terms of correlations by substituting the appropriate terms (as in Theorems 3 and 4).

longer be described as a diversity term; it necessarily contains accuracy terms. The desire to tease apart diversity and accuracy caused us to operate at the level of the signal.

4.2. Generality

Our framework is sufficiently general to be applied to a wide range of real-world situations. Depending upon the modeling environment, one need only adjust the relevant parameters to fit the model to the environment. For example, if an investigator has reason to believe that all forecasters use a prediction scale in a similar way, all variances could be set equal to one another. Alternatively, if one was able to correct for forecaster bias, the bias terms could be set to zero. This would result in forecaster variance and criterion covariance trading off with forecaster diversity for optimal group configuration. In this case, accuracy would take on a somewhat diminished role in determining the optimal group, although not as diminished as in the LP result (which would require setting criterion covariance equal to zero as well).

4.3. Alternative Weighting Approaches

Davis-Stober et al. (2010) consider optimal weighting schemes with respect to the standard linear model under the restriction that the weighting relationships among the variables (equivalently, forecasters) are chosen exogenously, as in Dawes' (1979) "improper linear model." The assumptions in Davis-Stober et al. (2010) differ from the present case in that they consider optimal weighting with respect to minimizing *maximal* mean squared error within the context of linear regression. Even so, Theorem 4 of Davis-Stober et al. (2010) provides optimal weighting, under these alternative assumptions, for the type coherence inter-forecaster correlation structure described here and in LP. These optimal weights differ somewhat from those presented in Theorem 2 because they are intended to be "conservatively" optimal. In other words, given uncertainty about the criterion variable being predicted, Davis-Stober et al. solve for the weighting scheme that minimizes the average squared error of the least favorable state of nature. Future work could consider other optimization criteria for determining optimal forecast group composition.

One way to further generalize our results would be to incorporate a fully Bayesian perspective. For example, we assume that forecaster biases are known, or at least reasonably well estimated. In many situations, knowing or having good estimates of the various parameter values may not be practical. For these situations, one could place prior densities over the various parameter values. Under such a framework, data collection would yield updated posterior values, which could be used to calculate optimal group configuration ratios. We leave such a formulation for future work.

4.4. Conclusion

We have presented a general framework for creating and weighting optimal forecasting groups. Our work has particular application to the formation of "small" forecasting groups. We find that determining the optimal group configuration depends upon several components, including the diversity of forecaster predictions, how the individual forecasts covary with the criterion variable, and forecaster biases. We identified the precise nature of how these different components trade off with one another when determining optimal group configuration and weighting.

Our model generalizes that of LP. The main result of LP was that (1) "For small groups, accuracy dominates, the group should consist primarily of the most accurate types" and (2) "For large groups, an opposite result holds. Within-type covariance, a proxy for model diversity, matters most." Our analysis confirms the first conclusion but not the second. The second conclusion only holds under the assumptions that (a) all forecasters are unbiased in their predictions and (b) the criterion being predicted is nonrandom. Once either or both of these restrictive assumptions are relaxed, we find that accuracy remains a critical component to forming optimal forecasting groups. We conclude that forming optimal forecasting groups depends critically on the accuracy-diversity trade-offs that we have identified.

Acknowledgments

The authors would like to thank Dr. Ilia Tsetlin for comments on an earlier draft.

Appendix A

Proof of Theorem 1. The result follows from direct calculation. See also Davis-Stober et al. (2014).

Proof of Theorem 2. We can write the optimal weighted average as the solution to the following equality constrained quadratic optimization problem:

$$\min f(\mathbf{w}) \quad \text{with } f(\mathbf{w}) = \mathbf{w}'(\boldsymbol{\Sigma}+\boldsymbol{\delta}\boldsymbol{\delta}')\mathbf{w} - \boldsymbol{\sigma}_{\mathbf{s},\mathbf{v}} + \mathrm{var}(V),$$

such that $\sum_{i=1}^{M} w_i = 1$. Note that the matrix $\boldsymbol{\Sigma}+\boldsymbol{\delta}\boldsymbol{\delta}'$ is a symmetric, positive semi-definite matrix as both Σ and $\boldsymbol{\delta}\boldsymbol{\delta}'$ are positive semi-definite and symmetric. Following standard methods (Boyd and Vandenberghe 2004) for solving equality constrained quadratic optimization systems, the set of all optimal weights $\mathbf{w}$ is the solution set to the following system of equations,

$$\begin{pmatrix} \boldsymbol{\Sigma}+\boldsymbol{\delta}\boldsymbol{\delta}' & \mathbf{u} \\ \mathbf{u}' & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{\mathbf{s},\mathbf{v}} \\ 1 \end{pmatrix};$$

the optimal set of weights is unique if, and only if, the matrix

$$\begin{pmatrix} \boldsymbol{\Sigma}+\boldsymbol{\delta}\boldsymbol{\delta}' & \mathbf{u} \\ \mathbf{u}' & 0 \end{pmatrix}$$

is nonsingular.

Proof of Theorem 3. Let $\mathrm{MSE}_{\mathrm{crowd}}$ be defined as the expected squared error of the crowd as follows:

$$\begin{aligned}\mathrm{MSE}_{\mathrm{crowd}} &= \frac{1}{M^2}\cdot\big(A(\delta_a^2+\mathrm{var}(a))+B(\delta_b^2+\mathrm{var}(b))+A(A-1)(\delta_a^2+\mathrm{cov}(a)) \\ &\quad +B(B-1)(\delta_b^2+\mathrm{cov}(b))+2AB(\delta_a\delta_b+\mathrm{cov}(a,b))\big) \\ &\quad -\frac{2(A\mathrm{cov}(a,V)+B\mathrm{cov}(b,V))}{M}+\mathrm{var}(V).\end{aligned}$$

The result follows by first fixing the total group size, M, and treating $\mathrm{MSE}_{\mathrm{crowd}}$ as a function of A. Hence, all B terms are equal to $M-A$. This yields

$$\begin{aligned}\mathrm{MSE}_{\mathrm{crowd}} &= \frac{1}{M^2}\cdot\big(A(\delta_a^2+\mathrm{var}(a))+(M-A)(\delta_b^2+\mathrm{var}(b)) \\ &\quad +A(A-1)(\delta_a^2+\mathrm{cov}(a))\big) \\ &\quad +\frac{1}{M^2}\cdot\big((M-A)(M-A-1)(\delta_b^2+\mathrm{cov}(b)) \\ &\quad +2A(M-A)(\delta_a\delta_b+\mathrm{cov}(a,b))\big) \\ &\quad -\frac{2(A\mathrm{cov}(a,V)+(M-A)\mathrm{cov}(b,V))}{M}+\mathrm{var}(V).\end{aligned}$$

To solve for the optimal proportion of type a forecasters, we take the first derivative of $\mathrm{MSE}_{\mathrm{crowd}}$ with respect to A, set it equal to zero, and solve the resulting equation for A. Doing so yields the unique solution

$$\begin{aligned}A &= \frac{[\mathrm{var}(b)-\mathrm{var}(a)]-[\mathrm{cov}(b)-\mathrm{cov}(a)]}{2(\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2)} \\ &\quad +(M([\mathrm{cov}(b)-\mathrm{cov}(a,b)]+[\delta_b^2-\delta_a\delta_b] \\ &\quad +[\mathrm{cov}(a,V)-\mathrm{cov}(b,V)]))\cdot(\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2)^{-1}.\end{aligned}$$

We solve for the resulting optimal proportion by dividing this value by M. This solution minimizes expected squared error if $(\delta_a-\delta_b)^2+\mathrm{TC}(a,b)>0$, following the standard second derivative test. This threshold generalizes the second order condition of Lamberson and Page (2012). Note that differences in bias can offset the type of coherence condition for determining whether this critical threshold of type a forecasters is a minimum or maximum. We note that the above optimal value of A is the unconstrained optimum. Hence, it is possible for the optimal proportion to be larger (smaller) than one (zero). For such cases, all optimal proportions larger (smaller) than one (zero) should be set equal to one (zero).

Proof of Theorem 4. The result follows from examining the optimal proportion of type a forecasters (Theorem 3) and letting $M\to\infty$. If

$$\begin{aligned}&f(M) \\ &= \frac{[\mathrm{var}(b)-\mathrm{var}(a)]-[\mathrm{cov}(b)-\mathrm{cov}(a)]}{2M(\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2)} \\ &\quad +\frac{[\mathrm{cov}(b)-\mathrm{cov}(a,b)]+[\delta_b^2-\delta_a\delta_b]+[\mathrm{cov}(a,V)-\mathrm{cov}(b,V)]}{\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2},\end{aligned}$$

then

$$\begin{aligned}&\lim_{M\to\infty} f(M) \\ &= \frac{[\mathrm{cov}(b)-\mathrm{cov}(a,b)]+[\delta_b^2-\delta_a\delta_b]+[\mathrm{cov}(a,V)-\mathrm{cov}(b,V)]}{\mathrm{TC}(a,b)+(\delta_a-\delta_b)^2}.\end{aligned}$$

Proof of Theorem 5. The result follows from setting the optimal proportion of type a forecasters obtained in Theorem 3 to less than $\frac{1}{2}$ and solving the inequality in terms of M.

Proof of Theorem 6. The result follows from setting the optimal proportion of type a forecasters obtained in Theorem 3 to less than (respectively, greater than) $1/M$ (respectively, $(M-1)/M$) and solving the respective inequalities.

Appendix B

Expressing the optimal fraction of type a forecasters as

$$F_a = \frac{[\Delta_{C_b}]+\delta_b[\Delta_B]+[\Delta_V]}{\Delta_{C_a}+\Delta_{C_b}+(\Delta_B)^2},$$

we compute the partial derivative of F_a with respect to each of the Δ variables. To simplify, we have substituted back for $\mathrm{TC}(\mathrm{a},\mathrm{b}) = \Delta_{C_a}+\Delta_{C_b}$.

The partial derivative of F_a with respect to Δ_{C_b} is equal to

$$\frac{\partial F_a}{\partial \Delta_{C_b}} = \frac{1}{\mathrm{TC}(\mathrm{a},\mathrm{b})+\Delta_B^2} - \frac{\Delta_{C_b}+\delta_b\Delta_B+\Delta_V}{(\mathrm{TC}(\mathrm{a},\mathrm{b})+\Delta_B^2)^2} = \frac{1-F_a}{\mathrm{TC}(\mathrm{a},\mathrm{b})+\Delta_B^2}.$$

The partial derivative of F_a with respect to Δ_{C_a} is equal to

$$\frac{\partial F_a}{\partial \Delta_{C_a}} = -\frac{\Delta_{C_b} + \delta_b \Delta_B + \Delta_V}{(\mathrm{TC}(a,b) + \Delta_B^2)^2} = \frac{-F_a}{\mathrm{TC}(a,b) + \Delta_B^2}.$$

The partial derivative of F_a with respect to Δ_B is equal to

$$\frac{\partial F_a}{\partial \Delta_B} = \frac{\delta_b}{\mathrm{TC}(a,b)+\Delta_B^2} - \frac{2\Delta_B(\Delta_C+\delta_b\Delta_B+\Delta_V)}{(\mathrm{TC}(a,b)+\Delta_B^2)^2} = \frac{\delta_b - 2\Delta_B F_a}{\mathrm{TC}(a,b)+\Delta_B^2}.$$

The partial derivative of F_a with respect to Δ_V is equal to

$$\frac{\partial F_a}{\partial \Delta_V} = \frac{1}{\mathrm{TC}(a,b) + \Delta_B^2}.$$

We can also take the partial derivative of F_a with respect to the bias of type b forecasters,

$$\frac{\partial F_a}{\partial \delta_b} = \frac{\Delta_B}{\mathrm{TC}(a,b) + \Delta_B^2}.$$

Proof of bias versus validity:

$$\begin{aligned}\frac{\partial F_a}{\partial \Delta_B} > \frac{\partial F_a}{\partial \Delta_V} \quad &\Rightarrow \quad \frac{\delta_b - 2\Delta_B F_a}{\mathrm{TC}(a,b) + \Delta_B^2} > \frac{1}{\mathrm{TC}(a,b) + \Delta_B^2},\\ &\Leftrightarrow \delta_b - 2\Delta_B F_a > 1,\\ &\Leftrightarrow F_a < \frac{\delta_b - 1}{2(\delta_b - \delta_a)} \quad \text{when } \delta_b > \delta_a.\\ &\Leftrightarrow F_a > \frac{\delta_b - 1}{2(\delta_b - \delta_a)} \quad \text{when } \delta_b < \delta_a.\end{aligned}$$

Proof of bias versus covariance:

$$\begin{aligned}\frac{\partial F_a}{\partial \Delta_B} > \frac{\partial F_a}{\partial \Delta_{C_b}} \quad &\Rightarrow \quad \frac{\delta_b - 2\Delta_B F_a}{\mathrm{TC}(a,b)+\Delta_B^2} > \frac{1-F}{\mathrm{TC}(a,b)+\Delta_B^2},\\ &\Leftrightarrow \delta_b - 2\Delta_B F_a > 1 - F,\\ &\Leftrightarrow F_a < \frac{\delta_b - 1}{2(\delta_b - \delta_a) - 1} \quad \text{when } 2(\delta_b - \delta_a) - 1 > 0.\end{aligned}$$

Proof of validity versus covariance:

$$\begin{aligned}\frac{\partial F_a}{\partial \Delta_V} > \frac{\partial F_a}{\partial \Delta_{C_b}} \quad &\Rightarrow \quad \frac{1}{\mathrm{TC}(a,b) + \Delta_B^2} > \frac{1 - F_a}{\mathrm{TC}(a,b) + \Delta_B^2}\\ &\Leftrightarrow 1 > 1 - F_a\\ &\Leftrightarrow F_a > 0.\end{aligned}$$

References

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).

Budescu DV, Chen E (2015) Identifying expertise to extract the wisdom of crowds. *Management Sci.* 62(2):267–280.

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *Internat. J. Forecasting* 5(4):559–609.

Cowgill B, Zitzewitz E (2015) Corporate prediction markets: Evidence from Google, Ford, and Firm X. *Rev. Econom. Stud.*, ePub ahead of print April 2, http://dx.doi.org/10.1093/restud/rdv014.

Davis-Stober CP, Dana J, Budescu DV (2010) A constrained linear estimator for multiple regression. *Psychometrika* 75(3):521–541.

Davis-Stober CP, Budescu DV, Dana J, Broomell SB (2014) When is a crowd wise? *Decision* 1(2):79–101.

Dawes RM (1979) The robust beauty of improper linear models in decision making. *Amer. Psych.* 34(7):571–582.

Fischer I, Budescu DV (1995) Desirability and hindsight biases in predicting results of a multi-party election. Caverni JP, Bar-Hillel M, Barron HF, Jungermann H, eds. *Contributions to Decision Research I* (North Holland, Amsterdam), 185–203.

Gigone D, Hastie R (1997) Proper analysis of the accuracy of group judgments. *Psychol. Bull.* 121(1):149–167.

Goldstein DG, McAfee RP, Suri S (2014) The wisdom of smaller, smarter crowds. *Proc. 15th ACM Conf. Electronic Commerce* (*EC'14*) (ACM, New York), 471–488.

Hong L, Page SE (2004) Groups of diverse problem-solvers can outperform groups of high-ability problem-solvers. *Proc. Natl. Acad. Sci.* 101(46):16385–16389.

Jose VRR, Grushka-Cockayne Y, Lichtendahl KC (2014) Trimmed opinion pools and the crowd's calibration problem. *Management Sci.* 60(2):463–475.

Lamberson PJ, Page SE (2012) Optimal forecasting groups. *Management Sci.* 58(4):805–810.

Mannes AE, Soll JB, Larrick RP (2014) The wisdom of select crowds. *J. Personality Soc. Psych.* 107(2):276–299.

Massey C, Simmons JP, Armor DA (2011) Hope over experience: Desirability and the persistence of optimism. *Psych. Sci.* 22(2):274–281.

Morgan GM, Adams PJ, Keith DW (2006) Elicitation of experts' judgments of aerosol forcing. *Climate Change* 75(1–2):195–214.

Simmons JP, Nelson LD, Galak J, Frederick S (2011) Intuitive biases in choice versus estimation: Implications for the wisdom of crowds. *J. Consumer Res.* 38(1):1–15.

Simon D, Holyoak KJ (2002) Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality Soc. Psych. Rev.* 6(4):283–294.

Sinclair TM, Joutz F, Stekler HO (2010) Can the fed predict the state of the economy? *Econom. Lett.* 108(1):28–32.

Surowiecki J (2004) *The Wisdom of Crowds* (Doubleday, New York).

Wallsten TS, Forsyth B, Budescu DV (1983) Stability and coherence of health experts' upper and lower subjective probabilities about dose-response curves. *Organ. Behav. Hum. Performance* 31(3):277–302.

Wallsten TS, Budescu DV, Erev I, Diederich A (1997) Evaluating and combining subjective probability estimates. *J. Behavioral Decision Making* 10(3):243–268.

Clintin P. Davis-Stober is an assistant professor of Psychological Sciences at the University of Missouri. He conducts research within the field of behavioral decision making. His research interests include the development of mathematical models of decision making, risk, and group prediction. He is a recipient of the William K. Estes Early Career Award from the Society for Mathematical Psychology.

David V. Budescu is the Anne Anastasi Professor of Psychometrics and Quantitative Psychology at Fordham University. His research is in the areas

of human judgment, individual and group decision making under uncertainty and with incomplete and vague information, and statistics for the behavioral and social sciences. He is past president of the Society for Judgment and Decision Making (2000–2001), fellow of the Association for Psychological Science, and an elected member of the Society of Multivariate Experimental Psychologists.

Stephen B. Broomell is an assistant professor of social and decision sciences at Carnegie Mellon University. He received his Ph.D. in quantitative psychology at the University of Illinois Urbana-Champaign and his B.S. in mathematics from the Ohio State University. His research interests include the study of judgment and decision making relating to risk perception, communication, understanding of scientific uncertainty, and information aggregation. He is also interested in methodology and modeling of forecasting and choice.

Jason Dana is an assistant professor of management and marketing at the Yale School of Management. He holds a Ph.D. from Carnegie Mellon University, was a postdoctoral fellow in quantitative psychology at the University of Illinois, and held a prior academic appointment at the University of Pennsylvania. His research and teaching interests lie in the areas of judgment and decision making and behavioral ethics.