Annals of Operations Research 23(1990)181–200

INVESTING IN THE STOCK MARKET: STATISTICAL POOLING OF INDIVIDUAL PREFERENCE JUDGMENTS

Noel CAPON
Graduate School of Business, Columbia University, New York, NY 10027, USA

and

Joel H. STECKEL
Department of Marketing, New York University, 40 West 4th Street, New York, NY 10003, USA

Abstract

Literature concerning the quality of individual and face-to-face group judgments has generally concluded that both groups and statistically pooled individuals outperform randomly chosen or average individuals. This paper extends previous research by comparing statistically pooled individual judgments of both individuals and face-to-face groups in a stock selection task. In general, decisions that would have resulted from statistically pooled judgments were better (as assessed by future stock value) than those that would have resulted from individual or face-to-face group judgments. In choosing among pooling methods, majority rule is often thought to be a very compelling criterion. However, majority rule can produce intransitive group preferences. Methods that use some procedure to resolve the intransitivities of majority rule did not perform well relative to other non-majority rule based methods. Another class of pooling methods, termed equity methods, used in conjunction with ordinal judgments, are recommended based on simplicity, performance, and fairness criteria. The results are discussed in terms of the nature of the task.

1. Introduction

Consider an investment trust committee deciding which security(ies) to puchase. How should members make the decision? Should they attempt group consensus? Should the judgment of an individual member with acknowledged expertise be accepted? Alternatively, should individual members form preferences *without* discussion and then combine or pool these preferences? And if yes, how?

A majority rule criterion provides a compelling method of combining judgments and resolving conflicts [8]. If there is an alternative that is preferred to every other by some majority, that alternative should be selected. The problem with this rationale is that intransitivities of majority rule often leave us without such an alternative. For example, consider three alternatives "α", "β", "γ" and three individuals "I", "II", "III". Suppose I's preference order is α, β, γ; II's is β, γ, α; and III's is γ, α, β. Application of the majority criterion produces α preferred to β, β preferred to γ, and γ preferred to α, a violation of transitivity.

© J.C. Baltzer AG, Scientific Publishing Company

A variety of methods have been propsed for resolving majority rule intransitivities (cf. refs. [3,9]). However, their rationales do not relate to which mechanisms might make the best decisions in practice. This statement also applied to majority rule in general. Perhaps then majority rule is not the best criterion. It is possible that other ways of combining individual preferences, or just letting the group decide the issue through face-to-face discussion, might be better.

This paper describes an experiment that compares the quality of group decisions with those that would have resulted from individual preference judgments and alternative methods of arithmetically combining or "statistically pooling" individual (possibly intransitive) preference judgments, in a common stock investment task. Subjects individually made preference judgments for a set of ten stocks. Then, assigned to groups and following face-to-face discussion, they made group judgments. From these data, we make the comparisons of interest; first for the entire sample, then for the portion that exhibited intransitive preferences.

There are two main reasons to examine statistical pooling methods for making judgments. First, they provide a baseline against which face-to-face group judgments can be compared [11]. An assessment of the role of social interaction in producing either a *process gain*, where the interaction improves the quality of a group's decision, or a *process loss*, where the interaction has a deleterious effect, can be made with respect to this baseline* [36]. Second, in the event of a process loss, the relative performance of different statistical-pooling methods may provide insight on how groups of individuals may improve their decision making. It is unlikely that a single method performs best uniformly, but a method or class of methods may be more robust than others.

The experiment reported in this paper differs from prior research in four important ways. First, rather than focus on either the individual versus face-to-face group, or individual versus statistical-pooling comparisons, we concentrate on the often neglected group versus statistical-pooling comparison [15]. Second, we investigate a more comprehensive set of statistical-pooling methods than the typically employed mean and median (see table 1). Third, we are interested in predictive judgments where subjects must form expectations of a future state. Finally, since judgment processes depend on the task [21,35], application of our results to organizational domains is achieved by choice of a business-related (stock market) judgment task.

In order to keep our task manageable, the pooling methods we investigate combine individual judgments in a *simple* arithmetic fashion. We examine neither those pooling methods that explicitly weight past performance (cf. ref. [26]), nor programmed procedures such as the Delphi [10] or Nominal Group techniques [39], both of which involve several stages of feedback. Neither do we examine the effects of context or individual differences (see ref. [20] for a review). Finally, we are primarily interested in the performance of the pooling methods employed. While systematic theoretical analyses of the similarities and differences among these methods would certainly lend

*Of course, investigation of a number of statistical-pooling methods provides several baselines.

Table 1

Summary of devices for statistically pooling individual preference judgments

hod	Description	Type of preference judgment
Dominance methods		
A. Copeland	N alternatives ranked according to the results of the $N(N-1)/2$ possible (binary) majority rule comparisons.	Preference rankings
B. Bowman–Colantoni	Find the ranking that minimizes the number of majority rule comparisons in which the winner is ranked below the loser.	Preference rankings
[. *Compromise methods*		
A. Weighted sums	Pooled preference judgment is weighted average of underlying individual judgments.	Cardinal utilities or preference rankings
1. Equal weights	All individuals weighted equally.	
2. Differential weights	Individuals weighted according to utility.	Cardinal utilities or preference rankings
B. Cook–Seiford	Find the ranking that minimizes the sum of the distances between itself and each of the individual judgments. Distance is defined to be the unique function that satisfies six reasonable axioms.	Preference rankings
III. *Equity methods*		
A. Geometric mean	Group utility is the geometric mean of the N individual utilities.	Cardinal utilities
B. Maximin rule	Maximize the minimum utility received by each individual.	Cardinal utilities or preference rankings

nsight as to why some methods might outperform others, that also is beyond the scope of this paper.

2. Background

2.1. JUDGMENTS OF MULTPLE INDIVIDUALS

Researchers have investigated the three possible pairs of two-way comparisons from face-to-face group, individual and statistically-pooled individual judgments, with varying degrees of effort. Face-to-face groups typically produce judgments that, with respect to some objective criterion, are superior to those made by the average or randomly chosen individual [25]. Face-to-face groups do not outperform the best member when best is defined *ex post* [25,36], i.e. after an objective measure of performance can be assessed. On the other hand, when the best member is identified *ex ante*, his/her performance is usually no better than that of the group [28,42]. Given that in actual decision making a best member would have to be identified *ex ante*, an effective best-member strategy becomes extremely diffficult to implement successfully.

Limited research comparing individual judgments with statistical pooling, typically the mean or median*, has usually shown pooling to be superior [17,37,43]. Hill's [20] review, comparing face-to-face groups with statistical pooling, suggests that although process losses dominate, process gains are occasionally found [22,42].

The state of extant research calls for further enquiry into the performance of statistically-pooled individual judgments. Whereas most prior research has concentrated on simple statistical-pooling techniques, other proposed methods have not been examined empirically. Dyer and Miles [13] did investigate several pooling techniques to recommend strategies for the 1977 Mariner Jupiter/Saturn project, but an objective standard to assess these methods did not exist.

2.2. JUDGMENT DOMAINS

Judgment domains can be classified according to a *type* dimension, evaluative or descriptive, and a *time* dimension, concurrent or predictive [30]. A judgment is evaluative if it is made on an affective criterion (e.g. good–bad, like–dislike). A judgment is descriptive if it is made on an objective criterion (e.g. dark–light, smart–dumb). The time dimension is self-explanatory. According to Castore [5], most group judgment research has been descriptive-concurrent, although recently some researchers have studied descriptive-predictive judgments (e.g. refs. [1,41]).

Stock *selection*, by contrast, represents an evaluation-predictive judgment. Slovic [33,34] focused specifically on the stock market as a key area for applied judgment research; Simon [32] identified financial decision-making as an important arena for testing behavioral phenomena. Other related studies include Clarkson's [6] research on trust investment officers, and Wright's [41] more recent study on the process and quality of individual judgments.

In descriptive judgment research, the performance criterion is the accuracy with which some *true* objective measure, such as expected returns, is estimated [15]. However, investment judges are concerned with making *better decisions*, and most models of investment behavior (cf. ref. [16]) incorporate a variety of considerations beyond expected return into preference functions for investments. Therefore, following Dyer and Miles [13], the judgments we examine are individual preferences or utilities.

*Ashton and Ashton [1] provide a recent exception by examining a weighted average of executives' advertising forecasts.

3. Methods of statistical pooling

In principle, a large number of statistical-pooling methods could be constructed and tested. We focus on a set of well-grounded methods to make our task manageable. Specifically, we test those used by Dyer and Miles [13], along with those of Bowman and Colantoni [3], Cook and Seiford [7], and Copeland [9]. These statistical-pooling methods fall into one of three categories: dominance, compromise, and equity methods (see table 1).

3.1. DOMINANCE METHODS

Dominance methods are those which appeal to the majority rule criterion but which use some procedure to resolve transitivity paradoxes.

Copeland's [9] method ranks N alternatives according to the $\frac{1}{2}N(N-1)$ possible majority-rule comparisons between all possible pairs of alternatives. The ranking is based on the number of $N-1$ alternatives each beats in a binary majority rule contest. For example, if α beats four alternatives and β beats three, α would be ranked higher that β even if β beat α. Ties are broken by comparing the number of losses and ties the alternatives had for the alternatives they did not beat.

Bowman and Colantoni [3] formulate an integer programming problem that focuses on paired comparisons. The integer program maximizes the number of paired comparisons consistent with the majority rule criterion, subject to constraints that ensure that group preferences are transitive.

Dominance methods can be justified only to the extent that a majority of the individuals are likely to make a "good" judgment. In one experiment, Holloman and Hendricks [22] found majority rule decisions to be inferior to actual group consensus. In addition, Huber and Delbecq [23] performed a simulation which demonstrated that majority rule judgments are inferior to simple averages of judgments in a variety of situations.

3.2. COMPROMISE METHODS

Compromise methods explicitly produce pooled judgments between the underlying individual judgments. The most common compromise methods belong to the class of *weighted sums* which follow the rule:

$$\text{Pooled preference judgment} = \frac{\sum_{i=1}^{N} w_i \times \text{individual } i\text{'s preference judgment}}{\sum_{i=1}^{N} w_i},$$

where the group size is N.

Differential weights (w_i) can be assigned to reflect perceptions of unequal abili ties of the individuals. For example, Dyer and Miles [13] set $w_i = 2.0$ for individual specializing in the type of data to be collected in the Mariner project and $w_i = 1.0$ fo those with more general interests. To the extent that specialization corresponds t ability, this scheme meets our objectives.

An important special case occurs when all w_i are equal. Recent research ha shown that equal weights outperform differential weights in many circumstances [12,14 Einhorn et al. (ref. [15], p. 160) note that:

> ...equal weights cannot reverse the relative weighting of the (individuals). For example, it is better to weight all group members equally than to assign high weights to those with poor judgment.

The individual judgments in weighted-sum methods can be either cardinal utili ties, such as those of von Neumann and Morgenstern [40], or preference ranking Rankings require relatively simple judgments, but strength of preference is obscured i the pooling process since no distinction is made between large and small difference between consecutively ranked alternatives. Furthermore, in the case of weighted sur of ranks, the numbers have no meaning beyond determining the group preferenc ordering. The simplest weighted-sum method, preference rankings and equal weight is associated with *Borda* [2], who proposed it to resolve elections in eighteenth-centu France.

Cook and Seiford's [7] compromise procedure relies on finding a consensu ranking, R^*, that minimizes the sum of the distances between R^* and the individu preference rankings,

$$R^* = \min_R \sum_{i=1}^{N} d(A^i, R),$$

where A^i denotes the preference ranking of the ith individual. The distance functi $d(A,B)$ is equal to $\sum_j |a_j - b_j|$, where $a_j(b_j)$ denotes the jth object's position in the over rank order $A(B)$. The distance function is derived from a set of "reasonable" axiom Cook and Seiford show that solving for R^* is equivalent to solving a linear programmi assignment problem.

3.3. EQUITY METHODS

Equity is a relevant concept if a group values the ongoing relationships amo its members and wants to keep them happy. The Nash [29] bargaining solution ma mizes the product of the increase in utility each individual receives with respect t "status quo", the utility each individual receives if an acceptable bargain cannot be stru The solution is a consequence of five axioms that Nash intended to characterize a f outcome. Dyer and Miles [13] assume the status quo is equivalent to losing all chan of gain, set this utility to zero, and form the following multiplicative (evaluation) ru

$$\text{Group utility} = \left[\prod_{i=1}^{N} (\text{individual's } i\text{'s utility})\right]^{1/N}.$$

The group utility is thus the *geometric mean* of the individual utilities; cardinal utilities are required for it to be meaningful.

The *maximin* (choice) rule, proposed by Rawls [31], maximizes the minimum utility received by each individual. It differs from most methods since the individual judgments must be preferences; they can, however, be either cardinal utilities or preference rankings.

4. Method

4.1. SUBJECTS

The subjects were fifty-six graduate students at the Graduate School of Business, Columbia University, randomly assigned to fourteen four-person groups. Although only a few subjects had professional investment experience, most had at least played the market. Each had completed at least one semester of study and taken at least one finance course.

4.2. STIMULI

The stimuli were ten pages from a recent Value Line investment survey, each providing extensive information regarding the ten-year historic financial performance, the nature of the products and markets, and Value Line's opinion regarding the short-term prospects for a publicly traded North American corporation. Seven corporations were listed on the New York Stock Exchange, one each on the American and Toronto stock exchanges, and one was traded over-the-counter. The corporations were randomly selected from that group of firms for which Value Line had published information sheets in the two weeks prior to the experiment. The ten corporations are listed in table 2, together with the number of shares that could be purchased for $1,000 on June 1st and the corresponding December 31st cash-out value.

4.3. PROCEDURE

Subjects were told that the study's purpose was to assess students' investment abilities. As a motivation to make their best efforts, the students were told that the best group would share a prize of $500. Each member of the fourteen groups was provided with the Value Line and current stock price information, and asked to study it individually for thirty minutes.

The students were told to imagine making a $1,000 investment in a *single* stock "today" (June 1st) that would be cashed out on December 31st. Fractional stocks could be purchased, transaction costs were zero, dividends declared would be reinvested at the prevailing stock price; should the company be acquired, all proceeds would be invested bility $(1 - p_i(m))$*. In the second lottery, the subject assigned a probability $p_j(m)$ for each stock j such that s/he was indiferent between having $1,000 invested in stock j for certain, and having it invested in his/her most preferred stock with probability $p_j(m)$ and an investment whose future value was zero with probability $(1 - p_j(m))$★.

The second lottery accounts for interpersonal differences in preferences *vis-à-vis* a sensible status quo and was used for the geometric mean (equity) method. The first lottery accounts only for differences between the ten stocks and was used for the weighted sums. Of course, the rank orders of the $p_i(m)$ and $p_j(m)$ for the two lotteries should be identical. Subjects were asked to check for inconsistencies and adjust their responses so that the rank orders corresponded.

When subjects completed both lotteries, they were told to reach group consensus on both first- and second-choice stock in which to invest $1,000. The second-choice stock was to be used *only if the first was unavailable*. These instructions were employed to further avoid portfolio considerations. Subjects were told that they had 20 minutes to complete this task and that a decision had to be made; otherwise they would forfeit their chance of winning the $500 prize.

After the selections, subjects completed questionnaires requiring them to allocate 100 points among the four group members (including themselves) in proportion to the *expertise* each had in the stock market. Following Dyer and Miles [13], individuals with point allocations in excess of 100 were assigned a weight $w_i = 2$, the remainder $w_i = 1$. Subjects also rated themselves on *experience*; how often s/he invested in the stock market according to a five-point scale: 1 = never, 5 = extensively. These scores generated an alternative set of weights; scores of 4 or 5 were assigned a weight $w_i = 2$, the remainder were assigned a weight $w_i = 1$. Both sets of differential weights were used for the differential weighted-sum methods.

4.4. ANALYSIS

The seven statistical-pooling methods (table 1), used in conjunction with cardinal utilities and preference rankings, generated 12 sets of rankings (see section A of tables 3 through 7). In addition, two best-member individual rankings, based on the best (*ex post*) actual performance and on the highest (*ex ante*) experience ratings, and the face-to-face group ranking were employed.

Four dependent variables were computed by comparing the actual ranking of the ten stocks' (12/31) values (table 2) to these rankings. Two variables focused on the choice of the single best stock: the actual 12/31 rank of the first-choice stock and a

*Indifference for individual i implies

Utility of stock $m = p_i(m) \times$ utility of most preferred $+ (1 - p_i(m)) \times$ utility of least preferred.

If we arbitrarily set the utility of the most preferred stock equal to 1 and the utility of the least preferred equal to zero, the utility of stock m is exactly equal to $p_i(m)$.

★An operation similar to that described in the above footnote, substituting the zero value option for the least preferred, yields a utility of $p_j(m)$ for stock m.

Table 2

Investment opportunities

	Number of shares purchased for $1,000 on 6/1	Value of $1,000 investment on 12/31	Final Rank	Standard Z-scores
Campbell Taggart[a]	43.48	1743.54	2	1.18
General Mills	24.39	1211.45	8	−0.39
Great A&P	163.93	1352.42	4	0.03
Holly Sugar	21.28	936.42	9	−1.20
Molson Cos.	40.00	1444.00	3	0.30
Norton Simon	52.63	1291.54	6	−0.15
Tyson Foods Inc.	55.56	1312.32	5	−0.09
United Cable TV	40.00	869.40	10	−1.56
Winn–Dixie	83.33	1254.27	7	−0.26
Zimmer Corporation	83.33	2008.25	1	1.97

[a]Campbell Taggart was purchased by Anheuser–Busch during the period of the study. The 12/31 value reflects the appropriate adjustments; see the method section.

in the acquiring company. The single stock nature of the task was employed to av portfolio considerations [27]. The task is a very real one. While normative finar theory suggests that investors take portfolio and covariance considerations into accou and indeed, professional investors do often behave that way, Clarkson's [6] study trust investment officers demonstrates that these investors do indeed evaluate sto without considering covariances in order to select a candidate list before making fi decisions (which would involve covariance considerations).

The single stock task was also necessary to mitigate against problems arisi from the nature of efficient capital markets. In an efficient market, all stocks are fai priced (assuming they are held as part of a well-diversified portfolio), and at the marg a rational investor would be indifferent between incremental investments in alternati stocks. However, if a stock is considered in isolation, or if the investment is r incremental (i.e. a discrete amount like $1,000), it no longer follows from effici market theory that a rational investor would be indifferent between alternative inve ments [16]. Indeed Wright [41] demonstrated that judges of stocks performed bet than randomly.

After thirty minutes, each subject completed two lotteries from which v Neumann–Morgenstern utilities could be obtained. In the first lottery, the subj assigned a probability $p_i(m)$ for each stock i such that s/he was indifferent betwe having $1,000 invested in stock i for certain, and having it invested in his/her m preferred stock with probability $p_i(m)$ and in his/her least preferred stock with prob

Value Line reports were helpful and subjects' experience facilitated better decisions. In any event, the judgments of some subjects appear more informed than suggested by a random walk view of stock prices.

5.2. STATISTICALLY-POOLED PREFERENCES AND PERFORMANCE

The nature of the data (ranks, small sample sizes, lack of independence, etc.) prevents formal statistical comparisons in most cases; nonetheless, some interesting patterns emerge by comparing tables 3 through 6.

First, average performance from statistical pooling across individuals within a group produces judgments superior to random individuals for each dependent variable, both for the entire sample and that portion which exhibits intransitive preferences.

Second, performance varies much more across groups than across methods. Formal two-way analyses of variance on the section A data in each of tables 3 through 6 produced four p-values for groups, each less than 0.001; the highest p-value for a method was 0.4299. These results are not surprising, since across-method (within group) comparisons use the same preference data for each analysis, whereas across-group (within method) comparisons do not. In addition, there is a natural high statistical correlation among many of the pooling methods. For example, if a sequence of random variables X_i are independent draws from a uniform distribution ranging from 0.0 to 1.0, the random variables $(1/10)\sum_{i=1}^{10} X_i$ and $[\prod_{i=1}^{10} X_i]^{1/10}$ have a correlation of 0.89 [13].

5.3. "BEST MEMBER", FACE-TO-FACE GROUP PREFERENCES, AND PERFORMANCE

Two common methods of identifying "best members" are *ex post*, after the results are known (rendering the method non-implementable), and the more relevant *ex ante*. As expected, the *ex post* best member outperformed all other judgments across all dependent variables in the present study, for both the whole sample and the intransitive subset. On the other hand, the performace of the *ex ante* best member approaches random; it is inferior to both the average individual and statistical pooling as well as the *ex post* best member. The implication is that it is difficult to identify the groups' best members.

Finally, consistent with Yetton and Bottger [42], the face-to-face group outperforms the *ex ante* best member. It performs approximately as well as the average individual (and maybe slightly worse). Nevertheless, each is inferior to *all* statistical pooling methods. All these findings apply both to the whole sample and the intransitive subset.

6. Discussion

This paper compares the performance of a variety of statistical-pooling methods, face-to-face group decisions, and individual preferences (average, *ex post* best member, *ex ante* best member). The overall superiority of statistical-pooling methods was

Actual ranking of first-choice stock

	Group														Means	
	1	2	3[a]	4	5	6	7	8	9	10	11	12	13	14	Total sample	Groups with intransitive preferences
A. *Statistical pooling*																
I. Dominance methods																
A. Copeland	2	9	7	2	2	7	2	5[b]	2	2	8	4.5[b]	2	1	4.04	4.25
B. Bowman–Colantoni	2	9	7	2	2	7	2	2	2	2	3	2	2	2	3.29	3.30
II. Compromise methods																
A. Weighted sums																
1. Ordinal preferences																
(a) Equal weights (Borda)	2	9	7	2	2	7	2	8	2	2	8	2	2	1	4.00	4.30
(b) Differential weights (experience)	2	7	7	2	2	2	2	8	10	2	8	2	2	2	4.14	4.50
(c) Differential weights (expertise)	2	9	N/A	2	2	7	2	8	2	2	8	2	2	1	3.77	4.30
2. Cardinal preferences																
(a) Equal weights	2	10	7	2	2	7	4[b]	2	2	2	8	7	2	1	4.14	4.30
(b) Differential weights (experience)	2	10	7	2	2	4.5[b]	4[b]	2	2	2	8	7	2	8	4.46	4.75
(c) Differential weights (expertise)	2	10	N/A	2	2	7	6	6[b]	2	2	2	7	2	1	3.85	4.10
B. Cook–Seiford	2	9	7	2	9	7	10	2	2	2	8	7	7	1	5.36	5.40
III. Equity methods																
A. Geometric mean (Nash)	2	7	7	2	2	7	2	2	2	2	3	2	2	8	3.57	3.70
B. Maximin rule																
1. Ordinal preferences	8	2	7	2	2	7	2	8	2	2	8	2	2	1	3.93	3.60
2. Cardinal preferences	8	10	7	2	2	1	9	8	2	2	8	2	2	8	5.07	4.50
B. *Actual judgment*																
Individuals																
I. *Ex post* best member	1	7	7	2	1	2	2	2	2	2	1	2	2	1	2.43	2.20
II. *Ex ante* best member	7	9	N/A	6	1	7	8	4	6	2	3	7	7	10	5.92	5.60
Face-to-face group decision	1	9	7	3	9	9	2	1	9	2	8	2	7	7	5.43	6.30
													Random value		5.50	

[a]N/A – Group 3 did not supply complete data.

[b][illegible]

Table 4
Z-scores of first-choice stock

	Group 1	2	3[a]	4	5	6	7	8	9	10	11	12	13	14	Means: Total sample	Means: Groups with intransitive preferences
A. *Statistical pooling*																
I. Dominance methods																
A. Copeland	1.18	−1.20	−0.26	1.18	1.18	−0.26	1.18	0.40[b]	1.18	1.18	−0.39	0.46b	1.18	1.97	0.66	0.57
B. Bowman–Colantoni	1.18	−1.20	−0.26	1.18	1.18	−0.26	1.18	1.18	1.18	1.18	0.30	1.18	1.18	1.18	0.74	0.71
II. Compromise methods																
A. Weighted sums																
1. Ordinal preferences																
(a) Equal weights (Borda)	1.18	−1.20	−0.26	1.18	1.18	−0.26	1.18	−0.39	1.18	1.18	−0.39	1.18	1.18	1.97	0.64	0.56
(b) Differential weights (experience)	1.18	−0.26	−0.26	1.18	1.18	1.18	1.18	−0.39	−1.56	1.18	−0.39	1.18	1.18	1.18	0.55	0.45
(c) Differential weights (expertise)	1.18	−1.20	N/A	1.18	1.18	−0.26	1.18	−0.39	1.18	1.18	−0.39	1.18	1.18	1.97	0.71	0.56
2. Cardinal preferences																
(a) Equal weights	1.18	−1.56	−0.26	1.18	1.18	−0.26	0.52[b]	1.18	1.18	1.18	−0.39	−0.26	1.18	1.97	0.57	0.54
(b) Differential weights (experience)	1.18	−1.56	−0.26	1.18	1.18	1.18	0.52[b]	1.18	1.18	1.18	−0.39	−0.26	1.18	−0.39	0.51	0.45
(c) Differential weights (expertise)	1.18	−1.56	N/A	1.18	1.18	−0.26	−0.15	0.40[b]	1.18	1.18	1.18	−0.26	1.18	1.97	0.65	0.62
B. Cook–Seiford	1.18	−1.20	−0.26	1.18	−1.20	−0.26	−1.56	1.18	1.18	1.18	−0.39	−0.26	−0.26	1.97	0.17	0.19
III. Equity methods																
A. Geometric mean (Nash)	1.18	−0.26	−0.26	1.18	1.18	−0.26	1.18	1.18	1.18	1.18	0.30	1.18	1.18	−0.39	0.78	0.65
B. Maximin rule																
1. Ordinal preferences	−0.39	1.18	−0.26	1.18	1.18	−0.26	1.18	−0.39	1.18	1.18	−0.39	1.18	1.18	1.97	0.69	0.80
2. Cardinal preferences	−0.39	−1.56	−0.26	1.18	1.18	1.97	−1.20	−0.39	1.18	1.18	−0.39	1.19	1.18	−0.39	0.32	0.51
B. *Actual judgments*																
Individuals																
I. *Ex post* best member	1.97	−0.26	−0.26	1.18	1.97	1.18	1.18	1.18	1.18	1.18	1.97	1.18	1.18	1.97	1.20	1.29
II. *Ex ante* best member	−0.26	1.20	N/A	−0.15	1.97	−0.26	−0.39	0.03	−0.15	1.18	0.30	−0.26	−0.26	−1.56	0.01	0.22
Face-to-face group decision	1.97	−1.20	−0.26	0.30	−1.20	−1.20	1.18	1.97	−1.20	1.18	−0.39	1.18	−0.26	−0.26	0.06	−0.14
Random value															0.00	

[a]N/A – Group 3 did not supply complete expertise data.
[b]Two or more stocks tied for first in the pooled order. The reported entry reflects their average.

Table 5
Actual rank sum of first- and second-choice stocks

	Group														Means	
	1	2	3[a]	4	5	6	7	8	9	10	11	12	13	14	Total sample	Groups with intransitive preferences
A. *Statistical pooling*																
I. Dominance methods																
A. Copeland	10	19	12[b]	10	9[b]	9	12	10	12	9	10	9	9	9	10.64	10.73
B. Bowman–Colantoni	10	16	15	10	9	8	12	10	12	9	11	9	9	3	10.21	9.91
II. Compromise methods																
A. Weighted sums																
1. Ordinal preferences																
(a) Equal weights (Borda)	10	19	9	10	7	8	8	10	12	9	10	9	9	9	9.93	10.09
(b) Differential weights (experience)	3	19	9	10	11	3	8	10	16	10	10	9	9	8	9.64	10.36
(c) Differential weights (expertise)	10	19	N/A	10	11	9	8	10	12	10	10	9	9	3	10.00	10.09
2. Cardinal preferences																
(a) Equal weights	10	19	9	10	9	8.5[b]	8	10	12	9	10	10	10	8.5[b]	10.21	10.36
(b) Differential weights (experience)	10	19	9	10	8	9	8	10	12	9	10	9	10	9	10.14	10.36
(c) Differential weights (expertise)	10	19	N/A	10	8	9	8	10	12	8.5[b]	10	9	10	9.5[b]	10.23	10.36
B. Cook–Seiford	3	14	9	10	14	15	12	9	8	9	10	9	9	3	9.57	10.27
III. Equity methods																
A. Geometric mean (Nash)	10	9	9	9	7	15	8	10	12	10	11	9	10	9	9.86	10.00
B. Maximin rule																
1. Ordinal preferences	10	12	9	10	10	9	5	10	12	9	10	10	10	8	9.57	9.55
2. Cardinal preferences	10	19	9	5	10	8	15	10	12	10	10	10	10	15	10.57	11.27
B. *Actual judgments*																
Individuals																
I. *Ex post* best member	5	9	9	5	8	6	12	6	8	9	3	9	7	3	7.07	6.91
II. *Ex ante* best member	9	14	N/A	14	10	15	15	5	15	10	11	9	9	2	10.62	10.91
Face-to-face group decision	4	19	9	5	12	13	11	11	11	12	11	12	9	9	10.57	11.18
												Random value			11.00	

[a] N/A – Group 3 did not supply complete expertise data.
[b] Two or more stocks tied for second in the pooled order. Their averages were used in computing the entries.

Table 6
Sum of Z-scores of first- and second-choice stocks

	Groups														Means	
	1	2	3[a]	4	5	6	7	8	9	10	11	12	13	14	Total sample	Groups with intransitive preferences
A. *Statistical pooling*																
I. Dominance methods																
A. Copeland	0.79	−2.76	0.18[b]	0.79	0.54[b]	0.92	−0.38	0.79	−0.38	0.92	0.79	0.92	0.92	1.67	0.41	0.20
B. Bowman–Colantoni	0.79	−1.46	−0.56	0.79	0.92	1.71	−0.38	0.79	−0.38	0.92	−0.09	0.92	0.92	3.15	0.57	0.63
II. Compromise methods																
A. Weighted sums																
1. Ordinal preferences																
(a) Equal weights (Borda)	0.79	−2.76	0.92	0.79	1.09	1.71	1.03	0.79	−0.38	0.92	0.79	0.92	0.92	1.67	0.66	0.60
(b) Differential weights (experience)	3.15	−2.76	0.92	0.79	−0.02	3.15	1.03	0.79	−1.71	0.79	0.79	0.92	0.92	1.67	0.75	0.49
(c) Differential weights (expertise)	0.79	−2.76	N/A	0.79	−0.02	0.92	1.03	0.79	−0.38	0.79	0.79	0.92	0.92	3.15	0.59	0.55
2. Cardinal preferences																
(a) Equal weights	0.79	−2.76	0.92	0.79	0.92	1.32[b]	1.03	0.79	−0.38	0.92	0.79	0.79	0.79	1.69[b]	0.60	0.54
(b) Differential weights (experience)	0.79	−2.76	0.92	0.79	1.03	0.92	1.03	0.79	−0.38	0.92	0.79	0.92	0.79	1.67	0.59	0.51
(c) Differential weights (expertise)	0.79	−2.76	N/A	0.79	1.03	0.92	1.03	0.79	−0.38	0.90[b]	0.79	0.92	0.79	1.06[b]	0.55	0.45
B. Cook–Seiford	3.15	−1.29	0.92	0.79	−1.29	−0.56	−0.38	0.92	1.03	0.92	0.79	0.92	0.92	3.15	0.71	0.14
III. Equity methods																
A. Geometric mean (Nash)	0.79	0.92	0.92	0.92	1.09	−0.56	1.03	0.79	−0.38	0.79	−0.09	0.92	0.79	1.67	0.69	0.68
B. Maximin rule																
1. Ordinal preferences	0.79	−0.38	0.92	0.79	0.79	0.92	1.48	0.79	−0.38	0.92	0.79	0.79	0.79	1.71	0.77	0.75
2. Cardinal preferences	0.79	−2.76	0.92	1.48	0.79	1.71	−1.59	0.79	−0.38	0.79	0.79	0.79	0.79	−0.56	0.31	0.17
B. *Actual judgments*																
Individuals																
I. *Ex post* best member	1.48	0.92	0.92	1.48	0.92	1.71	−0.38	2.00	1.03	0.92	3.15	0.92	1.09	3.15	1.38	1.45
II. *Ex ante* best member	0.92	−1.29	N/A	−0.54	0.77	−0.65	−1.35	2.00	−1.35	−0.79	−0.09	0.92	0.92	−0.38	−0.07	−0.25
Face-to-face group decision	2.27	−2.76	0.20	1.48	−0.90	−1.17	−0.02	0.41	−0.02	−0.38	−0.09	−0.38	0.92	0.92	0.09	−0.15
													Random value		0.00	

[a] N/A – Group 3 did not supply complete expertise data.

[b] Two or more stocks tied for second in the pooled order. Their averages were used in computing the entries.

expected from previous comparisons with individual preferences [15] and face-to-face groups [20], although there are exceptions [42]. A surprising finding was that face-to-face groups did not outperform average individuals. Although not unique (cf. refs. [4,18,38]), this finding is dwarfed by the amount of literature that supports the superiority of face-to-face groups over average individuals [20,25]. One salient question remaining is why face-to-face groups made such relatively poor decisions. Apparently, either the information exchanged was incorrect or "groupthink" [24] took over, and, in a search for cohesion, the group members paid more attention to personal factors than to the task at hand.

Steiner [36] identifies three critical factors in determining how well groups perform specific tasks: the type of task and its demands, the resources (here investment expertise) at the group's disposal, and the process used by the group. Since the individuals had adequate task resources, as evidenced by the statistical-pooling results, and since the process (face-to-face discussion) was similar to previous research, the difference in results from prior studies is likely due to the task.

Hackman and Morris [19] discuss the common view that groups outperform individuals (in part) because interaction among group members helps catch and remedy errors that individuals might not identify. This effect is most likely to occur in descriptive-concurrent domains [30], where an objectively correct answer exists, and there is an immediacy that lends itself to rational evaluation and consequent error corrections. As noted earlier, most research on group judgment falls into this category.

By contrast, a stock selection task requires evaluative-predictive judgments. These judgments are unlikely to elicit the error correction mechanism, since assumptions underlying the predictions are likely to be highly idiosyncratic. Face-to-face groups would not, therefore, be expected to outperform average individuals.

Furthermore, the idiosyncratic nature of the assumptions makes it difficult for individuals to identify the best member, and tends to produce similar, and poor, face-to-face group and *ex ante* best member performance. Miner [28] found that groups could not identify their best member, although Ashton and Ashton [1] developed a weighting scheme based on ongoing relationships that worked quite well.

We found no significant differences among pooling methods. We may, nonetheless, seek relative performance by ranking the methods (table 7), although the lack of significant differences renders the interpretations quite tentative but useful perhaps as future hypotheses. Table 7 suggests that for most of the pooling techniques that can employ both ordinal and cardinal preferences, pooling (the simpler) ordinal preferences outperforms pooling cardinal preferences. Apparently, the extra strength of preference information extracted in moving from ordinal to cardinal preferences was likely to be incorrect. Use of ordinal preferences guards against this type of mistake in the same manner that equal weighting of individuals cannot reverse their relative expertise.

Additionally, no clear pattern emerges when the dependent variable shifts from single-stock to two-stock choice (see table 7). The relative performance of some pooling techniques improves (e.g. differential weights-experience and maximin, both with preference rankings, Cook–Seiford); others regress (e.g. differential weights-expertise

Table 7

Rank of average performance for five dependent variables

	Actual rank of first-choice stock (table 3)	Z-score of first-choice stock (table 4)	Actual rank sum of first- and second-choice stock (table 5)	Sum of Z-scores of first- and second-choice stocks (table 6)
A. *Statistical pooling*				
I. Dominance methods				
A. Copeland	8 (6)[a]	6 (6)	15 (12)	12 (12)
B. Bowman–Colantoni	2 (2)	3 (3)	9 (3)	10 (4)
II. Compromise methods				
A. Weighted sums				
1. Ordinal preferences				
(a) Equal weights (Borda)	7 (7)	8 (7)	6 (5)	6 (5)
(b) Differential weights (experience)	9 (10)	10 (11)	4 (8)	3 (9)
(c) Differential weights (expertise)	4 (7)	4 (7)	7 (5)	8 (6)
2. Cardinal preferences				
(a) Equal weights	9 (7)	9 (9)	9 (8)	7 (7)
(b) Differential weights (experience)	11 (12)	11 (11)	8 (8)	8 (8)
(c) Differential weights (expertise)	5 (5)	7 (5)	11 (8)	11 (10)
B. Cook–Seiford	13 (13)	13 (14)	2 (7)	4 (10)
III. Equity methods.				
A. Geometric mean (Nash)	3 (4)	2 (4)	5 (4)	5 (3)
B. Maximin rule				
1. Ordinal preferences	6 (3)	5 (2)	2 (2)	2 (2)
2. Cardinal preferences	12 (10)	12 (10)	12 (13)	13 (13)
B. *Actual judgments*				
Individuals				
I. *Ex post* best member	1 (1)	1 (1)	1 (1)	1 (1)
II. *Ex ante* best member	15 (14)	15 (13)	14 (14)	15 (15)
Face-to-face group decision	14 (15)	14 (15)	12 (15)	14 (14)

[a] Values in parentheses apply to the groups that exhibit intransitive preferences.

[4] J.P. Campbell, Individual versus group problem solving in an industrial sample, J. Appl. Psychol. 52(1968)106–112.

[5] C. Castore, Decision making and decision implementation in groups and organizations, in: *Managerial Control and Organizational Democracy*, ed. B. King, S. Streufert and F. Feidler (Wiley, New York, 1978) pp. 267–275.

[6] G.P.E. Clarkson, *Portfolio Selection: A Simulation of Trust Investment* (Prentice-Hall, Englewood Cliffs, NJ, 1962).

[7] W.D. Cook and L.M. Seiford, Priority ranking and consensus formulation, Manag. Sci. 24(1978)1721–1732.

[8] Condorcet, Marquis de, *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix* (Paris, 1785).

[9] A.H. Copeland, A "reasonable" social welfare function, University of Michigan Seminar on Applications of Mathematics to the Social Sciences (1951).

[10] N.C. Dalkey, The Delphi method: An experimental study of group opinion, RM-5888-PR, The Rand Corporation, Santa Monica, CA (1969).

[11] J.H. Davis, *Group Performance* (Addison–Wesley, Reading, Mass., 1969).

[12] R.M. Dawes and B. Corrigan, Linear models in decision making, Psychol. Bull. 81(1974)95–106.

[13] J.S. Dyer and R.F. Miles, Jr., An actual application of collective choice theory to the selection of trajectories for the Mariner Jupiter/Saturn 1977 project, Oper. Res. 21(1976)220–244.

[14] H.J. Einhorn and R.M. Hogarth, Unit weighting schemes for decision making, Organizational Behavior and Human Performance 13(1975)171–192.

[15] H.J. Einhorn, R.M. Hogarth and E. Klempner, Quality of group judgments, Psychol. Bull. 84(1977)158–172.

[16] E.F. Fama and M.H. Miller, *The Theory of Finance* (Dryden Press, Hinsdale, Ill., 1972).

[17] K. Gordon, A study of esthetic judgments, J. Exp. Psychol. 6(1923)36–43.

[18] W.K. Graham, Acceptance of ideas generated through individual and group brainstorming, J. Soc. Psychol. 101(1977)231–234.

[19] J.R. Hackman and C.G. Morris, Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration, in: *Advances in Experimental Social Psychology*, ed. L. Berkowitz (Academic Press, New York, 1975), Vol. 8, pp. 45–99.

[20] G.W. Hill, Group versus individual performance: Are $N+1$ heads better than one?, Psychol. Bull. 91(1982)517–539.

[21] R.M. Hogarth, Cognitive processes and the assessment of subjective probability distributions, J. Amer. Statist. Assoc. 70(1975)271–294.

[22] C.R. Holloman and H.W. Hendrick, Adequacy of group decisions as a function of the decision making process, Academy of Management J. 15(1972)175–184.

[23] G.P. Huber and A. Delbecq, Guidelines for combining the judgments of individual members in decision conferences, Academy of Management J. 15(1972)161–174.

[24] I.L. Janis, *Victims of Groupthink* (Houghton-Mifflin, Boston, 1972).

[25] I. Lorge, D. Fox, J. Davitz and M. Brenner, A survey of studies contrasting the quality of group performance and individual performance, 1920–1957, Psychol. Bull. 55(1958)337–372.

[26] S. Makridakis and R.L. Winkler, Averages of forecasts: Some empirical results, Manag. Sci. 29(1983) 987–996.

[27] H. Markowitz, *Portfolio Selection* (Wiley, New York, 1959).

[28] F.C. Miner, Jr., Group versus individual decision making: An investigation of performance measures, decision strategies, and process losses/gains, Organizational Behavior and Human Performance 33(1984)112–124.

[29] J.F. Nash, Jr., The bargaining problem, Econometrica 13(1950)155–162.

[30] J.C. Naylor, R.D. Pritchard and D.R. Ilgen, *A Theory of Behavior in Organizations* (Academic Press, New York, 1980).

with cardinal preferences, Bowman–Colantoni, Copeland, geometric mean); and others remain essentially the same (e.g. equal weights with both cardinal utilities and preference rankings, maximin with cardinal utilities). Similarly, no clear pattern emerges when we restrict the analysis to groups with intransitive preferences except that the advantage of ordinal preferences over cardinal preferences disappears for weighted sums and single-stock dependent variables.

Finally, the two equity methods perform relatively well regardless of whether one or two stocks are selected, whether the whole sample or the intransitive subset is used. The geometric mean method produces ranks 3, 2, 5, 5 for the four dependent variables; maximin rule with preference rankings produces 6, 5, 2, and 2 for the whole sample. Even if one suspects that these differences are indeed random, the equity methods are attractive since equity is, in general, a desirable objective.

The performance of two equity methods is particularly important from a prescriptive standpoint; they outperformed the single pooling techniques upon which most prior research is based. A stock is rated highly by these equity methods if, and only if, it is rated highly by each individual. Compared to the compensatory compromise and dominance methods, these are non-compensatory procedures. To the extent that errors are difficult to find and best members are difficult to identify, equity methods maximize the likelihood that *(ex post)* best members have a large influence on the outcome. In addition, equity methods should minimize conflict and post-decision remorse among group members and thus produce outcomes that are acceptable to all members.

It is tempting to conclude that collective predictive judgments should be made by statistically aggregating individual judgments, in particular by use of equity methods. However, the old adage "two heads are better than one" has much intuitive appeal. Perhaps the ultimate challenge is to develop mechanisms by which individual members can work together constructively so that genuinely creative outcomes may result (ref. [19], p. 46).

Acknowledgements

The authors acknowledge the assistance of Kim P. Corfman and Barbara A. Welch. They are both grateful for financial support from the Faculty Research Fund at the Graduate School of Business, Columbia University, and the Redward Foundation.

References

[1] A.H. Ashton and R.H. Ashton, Aggregating subjective forecasts: Some empirical results, Manag. Sci. 31(1985)1499–1508.

[2] J.-D. de Borda, *Mémoire sur les Elections au Scrutin: Histoire de L'Academie Royale des Sciences* (1781).

[3] V.J. Bowman and C.S. Colantoni, Majority rule under transitivity constraints, Manag. Sci. 19(1973) 1029–1041.

[31] J. Rawls, *A Theory of Justice* (Belknap Press of Harvard University Press, Cambridge, Mass., 1971).
[32] H. Simon, *The New Science of Management Decision* (Harper, New York, 1960).
[33] P. Slovic, Analyzing the expert judge: A descriptive study of a stockbroker's decision process, J. Appl. Psychol. 53(1969)255–263.
[34] P. Slovic, Psychological study of human judgment: Implications for investment decision making, J. Finance 27(1972)779–799.
[35] P. Slovic and S. Lichtenstein, Comparison of Bayesian and regression approaches to the study of information processing in judgment, Organizational Behavior and Human Performance 6(1971)649–744.
[36] I.D. Steiner, *Group Process and Productivity* (Academic Press, New York, 1972).
[37] J.B. Stroop, Is the judgment of the group better than that of the average member of the group?, J. Exp. Psychol. 15(1932)550–562.
[38] D.W. Taylor, P.C. Berry and C.H. Block, Does group participation when using brainstorming facilitate or inhibit creative thinking?, Admin. Sci. Quart. 3(1958)23–47.
[39] A.H. van de Ven and A.L. Delbecq, The effectiveness of nominal, Delphi, and interacting group decision making processes, J. Academy of Management 17(1974)605–621.
[40] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 2nd ed. (Princeton University Press, Princeton, NJ, 1947).
[41] W.F. Wright, Properties of judgment models in a financial setting, Organizational Behavior and Human Performance 23(1979)73–85.
[42] P.W. Yetton and P.C. Bottger, Individual versus group problem solving: An empirical test of a best member strategy, Organizational Behavior and Human Performance 29(1982)307–321.
[43] R.B. Zajonc, A note on group judgments and group size, Human Relations 15(1962)177–180.