Published in Journal of Applied Psychology (2024) 109(4), 513-533. DOI: 10.1037/apl0001141

Designing Pareto-Optimal Selection Systems for Multiple Minority Subgroups and Multiple Criteria

Wilfried De Corte[1], Paul R. Sackett[2], and Filip Lievens[3]
[1] Faculty of Psychology, Department of Data-Analysis, Ghent University
[2] Department of Psychology, University of Minnesota
[3] Lee Kong Chian School of Business, Singapore Management University

Currently used Pareto-optimal (PO) approaches for balancing diversity and validity goals in selection can deal only with one minority group and one criterion. These are key limitations because the workplace and society at large are getting increasingly diverse and because selection system designers often have interest in multiple criteria. Therefore, the article extends existing methods for designing PO selection systems to situations involving multiple criteria and multiple minority groups (i.e., multiobjective PO selection systems). We first present a hybrid multiobjective PO approach for computing selection systems that are PO with respect to (a) a set of quality objectives (i.e., criteria) and (b) a set of diversity objectives where each diversity objective relates to a different minority group. Next, we propose three two-dimensional subspace procedures that aid selection designers in choosing between the PO systems in case of a high number of quality and diversity objectives. We illustrate our novel multiobjective PO approaches via several example applications, thereby demonstrating that they are the first to reveal the complete gamut of eligible PO selection designs and to faithfully capture the Pareto trade-off front in case of more than two objectives. In addition, a small-scale cross-validation study confirms that the resulting PO selection designs retain an advantage over alternative designs when applied in new validation samples. Finally, the article provides a link to an executable code to perform the new multiobjective PO approaches.

Keywords: adverse impact, personnel selection, Pareto-optimal, selection design, multiobjective optimization

Supplemental materials: https://doi.org/10.1037/apl0001145.supp

The past few years have witnessed a growing interest in Pareto-optimal (PO) selection system design as a viable strategy for developing selection systems that aim to reconcile the often conflicting goals of selection quality (e.g., the expected job performance of the selected applicants) and selection diversity (i.e., an adequate representation of the minority candidates among the selected applicants). In a recent article, Rupp et al. (2020) listed over 30 papers, presentations, and dissertations devoted to the subject of the PO approach for designing biobjective selection systems (i.e., selection systems for situations where both the goals of quality, with respect to a single criterion, and diversity, with respect to a single minority group, are valued). The approach merits this interest because it results in selection systems that offer a PO trade-off of the selection objectives, meaning that each of these systems is expected to result in values for the objectives such that any improvement in one objective can only take place if at least one other objective worsens.

Despite the growing interest in the PO approach, its actual usage in practice is still lagging behind. On the basis of conversations with senior executives of some of the top employers of I-O psychologists in the United States, we identified three key obstacles that seem to impede a wider adoption of the PO method. The three obstacles are (a) the mathematical sophistication behind the method, (b) the concern that the PO results are overly optimistic when applied to new situations, and (c) the limitation of the presently available PO method to situations involving only a single minority group. Rupp et al. (2020) offered a solution for the first obstacle. They provided not only detailed checklists and flow charts of key steps to follow when implementing the PO approach in practice, but also presented and illustrated a simple and easy to use tool (i.e., the "ParetoR Shiny app," developed by Song et al., 2017) to execute the PO approach. In addition, various studies (De Corte et al., 2020, 2022; Song et al., 2017; Wee et al., 2014) conducted research related to the second concern about the cross-validity potential of calibration PO selection systems when applied in a new validation setting. For example, De Corte et al. (2022) compared the three main types of cross-validity[1] of both PO and fixed weight (FW) selection systems. PO systems outperformed FW systems for applicant pool sizes as small as 100,

This article was published Online First October 26, 2023.
Wilfried De Corte https://orcid.org/0000-0001-7400-3181
The computational resources and services used in this work were provided to Wilfried De Corte by the Vlaams Supercomputer Centrum (Flemish Supercomputer Center), funded by the Research Foundation–Flanders (FWO) and the Flemish Government–Department EWI Economie Wetenschap en Innovatie. The authors have no known conflict of interest to disclose.
Correspondence concerning this article should be addressed to Wilfried De Corte, Faculty of Psychology, Department of Data-Analysis, Ghent University, Bosstraat 6, 9080 Lochristi, Belgium. Email: wilfried.decorte@ugent.be

[1] The three studied cross-validity conditions are (a) population-to-sample cross-validity, investigating the cross-validity of PO systems developed from population data when applied to an applicant sample; (b) sample-to-population cross-validity, investigating the cross-validity of PO systems developed from sample data when applied to the applicant population; (c) sample-to-sample cross-validity, focusing on the cross-validity of PO systems developed from applicant sample data when applied to a new applicant sample.

even in the practically most important sample-to-sample cross-validity context.

As compared to the first two obstacles, the third obstacle, namely the limitation of available methods for designing PO selection systems to situations involving only a single minority group, has still remained largely unresolved. As shown in our review of previous research below, no generally applicable procedure for designing *multiobjective* PO selection systems (i.e., single and multistage selection systems that are PO with respect to more than two-valued objectives, including situations with multiple quality objectives and/or multiple minority groups) is presently available. Yet, in today's diverse society applicant groups often count members of different minority groups challenging employers to devise selection systems that, conditional on the available predictors/criterion information, are expected to be PO with respect to the pursued quality goal as well as to the representation of all present minority groups (e.g., Black, Hispanic, and female minority groups). Unfortunately, and even though researchers and practitioners have already up-to-date meta-analytic information about the subgroup differences of selection procedures related to minority groups (i.e., Hispanics and Asians) other than Blacks (Roth et al., 2017), the PO approach (De Corte et al., 2007; Rupp et al., 2020) does not permit them to capitalize on this information. Given the limitation of use with only one minority group at the time, even the repeated usage of the PO approach, each time focusing on a different subset of two of the objectives, at best results in a very partial retrieval of the selection systems that are PO with respect to the entire set of pursued objectives.[2]

While it is concern about multiple subgroups that drove our initial interest in developing multiobjective PO systems, we also note that it is not uncommon for selection system designers to have interests in the prediction of multiple criteria, such as task performance, organizational citizenship, and counterproductive work behavior. Thus, we broadened our focus to multiobjective PO systems involving multiple subgroups, multiple criteria, or both.

Therefore, the first key purpose of the article is the development of an efficient procedure to derive PO selection systems and the corresponding front of PO trade-offs[3] for single and multistage selections where both several different quality and several different diversity goals are valued. In particular, the procedure aims to obtain a list of PO systems that together provide a high quality, discrete representation of the entire PO front for these selection situations and where each PO system in the list is characterized in terms of both the corresponding trade-off value (i.e., the set of values for the valued objectives) and the corresponding set of design parameter values (see the section "Designing PO Selection Systems"). Unless otherwise stated, the quality objectives are operationalized as (i.e., equated to) the expected performance of the selected applicants on the different quality goals, whereas the adverse impact ratio (AIR, defined as the ratio between the hiring rate in the minority group as compared to the hiring rate in the majority group) of the selection in the different minority groups gauges the diversity goals. The latter choice is identical to the operationalization of the diversity goal in most of the research on biobjective PO selection system (e.g., De Corte et al., 2011; Rupp et al., 2020), and, together with the former choice, it has the unique advantage that the operationalization can be used in designing both single and multistage selections.

Choosing between alternative PO systems becomes considerably more complicated in situations with an increasing number of minority applicant groups and/or quality outcomes. Whereas the resulting PO quality/diversity trade-offs of the biobjective PO approach can be displayed as a curve segment in the quality/diversity plane and summarized in a table containing a representative subset of only two-valued quality/diversity PO trade-offs, the situation involving three objectives (e.g., a single quality and two diversity objectives, diversity1 and diversity2, for example) leads to a surface in the three-dimensional quality/diversity1/diversity2 space and to a table where each entry has already three elements. With four or more objectives a visual representation of the PO trade-off hypersurface is even impossible and the corresponding tabular summary becomes increasingly unclear, making it very difficult for selection designers to evaluate and compare the merits of the different PO alternatives in order to make a final choice.

The second key contribution of the article addresses the above choice issue for situations involving one or more quality objectives and at least two diversity objectives by presenting three additional procedures. Each of these procedures solves for a different set of selection systems that are not only PO with respect to an aggregated quality objective and an aggregated diversity objective, but also PO with respect to each of the individual quality and diversity objectives. As a consequence, the resulting solutions cannot only be represented as part of the entire, full dimensional PO trade-off hypersurface but can also, and similar to the solutions of the biobjective PO approach, be depicted in a two-dimensional space. We henceforth refer to these additional procedures as the *two-dimensional subspace* procedures.

As a third and final contribution, we investigate via simulation studies whether designs derived by the new multiobjective PO approach still outperform alternative designs, such as fixed weight (FW) and expert weight (EW) systems, under cross-validation. Similar to biobjective PO methods, the new hybrid and the subspace procedures result in selection designs that are PO when applied to the applicant *population*: The designs are PO in terms of the *expected* trade-off of the selection objectives.[4] In applicant samples, the designs may perform less well, especially in the typical case that the computation of the systems is based on sample rather than on population data about the predictor/criterion effect sizes and intercorrelations.

As an introduction, we first recapitulate the basics of PO selection system design in general and discuss previous research on multiobjective PO selection design in the next two sections.

Designing PO Selection Systems

De Corte et al. (2007, 2011) introduced the PO approach to selection system design for situations with a mixture of majority and minority applicants and where both the goals of selection quality and diversity are valued. The central tenet of the approach is that in these

[2] The example discussed in the section "Comparison with repeated application of the biobjective PO approach" illustrates this.

[3] See the section "Designing PO Selection Systems" for a definition of the terms PO front and PO trade-offs.

[4] Existing PO approaches use the population formula to compute the value of the selection objectives as a function of the selection design parameter values. Except when using the predictor composite validity and effect size to gauge the selection quality and diversity objective, the analytic computation of selection designs that are PO in finite sized applicant pools is not possible (cf. De Corte et al., 2022).

situations the only acceptable designs are those that (conditional on the available predictors/criterion information) are expected to result in a PO trade-off of the valued objectives. All other designs are judged unacceptable because they are expected to result in a quality and diversity trade-off that is *dominated* by the expected trade-off of a PO system, meaning that at least one of the *trade-off components* (i.e., selection quality or diversity) of the system has a smaller value than the value of the corresponding trade-off component of the PO system, whereas the other component values are at most as high as the corresponding PO trade-off component values.

As PO selection systems aim to maximize multiple interdependent objectives, the determination of these systems is formally equivalent to that of solving a multiobjective optimization problem (MOOP). For more than half a century, the problem and the ways to solve it have received a lot of attention in the optimization literature (see, e.g., the entry "Multiobjective optimization" in Wikipedia). The generic format of a MOOP comprises (a) the different objectives that are to be optimized, often referred to as the *objective functions*, (b) the parameters or problem variables on which the value of the objective functions depends (the *design parameters* or *design variables*), and (c) a number of *constraint functions* that demarcate the set of admissible design parameter values. In PO selection design, the objective functions correspond to the quality and diversity objectives; the design parameters are the weights with which the selection predictors are combined to selection composites and the selection rates applied in the selection; and the constraint functions typically impose bounds on the predictor weights (e.g., only nonnegative weights) and the selection rates at completion of the intermediate selection stages (henceforth also referred to as retention rates), but may also express restrictions on the staging, the cost and the timing of the selection.

When the importance of the objectives is not specified in advance,[5] as is the case in the PO approach to selection system design, a MOOP does not have a single solution but rather an infinite number of solutions, that are collectively referred to as the PO front, and where each solution corresponds to a particular PO trade-off, that is, to a particular combination of values for the objectives. The methods for solving the MOOP typically aim for a discrete approximation or representation of the PO front, and the criteria of coverage (i.e., the extent to which the entire PO front is captured), uniformity (all parts of the front are fairly equally represented) and cardinality (the number of PO solutions obtained by the method) are often used to evaluate the quality of the solution set obtained by the method (cf. Sayin, 2000).

There are roughly two kinds of methods for solving a MOOP when the preference or importance of the objectives is undetermined: analytic methods based on mathematical programming and making use of differential calculus (i.e., gradient and Hessian information) and methods based on a metaheuristic principle such as the evolutionary process, the swarming of ant colonies and so on. The former methods are applicable when the objective and constraint functions of the MOOP are (twice) continuously differentiable in the design parameters, whereas the metaheuristic methods impose no such requirements. When applicable, as is true in case of the derivation of PO selection system designs, analytic-based methods are the preferred option because they result in a much higher quality discrete representation of the PO front than achieved by the metaheuristic-based methods.

The procedure presented here for deriving selection systems that are PO with respect to both multiple quality and diversity objectives (cf. the first key contribution of the article) is based on three recently proposed analytic methods for solving the MOOP: the normal boundary intersection (NBI) method of Das and Dennis (1998), the enhanced normalized normal constraint (ENNC) method of Messac and Mattson (2004), and the successive boundary generation (SBG) method of Mueller-Gritschneder et al. (2009). De Corte et al. (2007) details the NBI method, whereas the other methods are discussed in Appendix A. Here, we note that the new hybrid multiobjective PO procedure, as compared to the existing NBI, ENNC, and SBG methods, has the advantage of resulting in a full and quite even discrete representation of the PO front, thereby satisfying the coverage and uniformity criteria of a high-quality solution, whereas the cardinality of the representation can be controlled.

PO Selection Design and Dealing With Multiple Minority Groups: Previous Research

So far, previous research on the design of selection systems that are PO with respect to multiple minority groups has tackled the issue from three different angles. The first is represented by Peterson and Morris (2021) who explored the fit between PO systems, as developed for one minority group, when applied to a second minority group. Initial results show that PO solutions developed separately for Black and Hispanic minority groups are fairly similar, thereby mitigating concerns that use of PO systems focused on one minority group will inadvertently discriminate against another. Obviously, this result depends on the similarity of the selection predictor effect sizes[6] in the two minority groups and will no longer hold when the groups show markedly different effect size values.

Andrews and Geden (2021) exemplified the second angle of attack. They used different evolutionary multiobjective optimization (EMOO) methods for developing multiobjective PO selection systems and compared the merit of these systems to that of a rational weighting system when all are applied to a hold-out sample of applicants. On the basis of their example data, they concluded that the PO systems consistently outperform the system based on a rational weighting of the selection predictors. They also found a considerable degradation in the quality of the discrete approximation of the PO front with an increasing number of objectives (i.e., from four objectives onwards). However, the proposal can be applied only to the design of single stage selection systems where the validity of the selection predictor composite is used to gauge the quality objective. Also, although EMOO methods may offer a valuable alternative to the computation of the PO systems, they are

[5] In discussing the MOOP, one (e.g., Nagy et al., 2020), often distinguishes between situations where (a) information on the preference, importance or weight of the different objectives is available, the so-called a priori situation, (b) such information comes progressively available, the so-called interactive situation, and (c) no such information is available and the decision on the preferred PO solution is postponed until the full PO front has been derived, the a posteriori situation. The former as well as the present PO approach fall in the latter category, where the preference is undetermined as no preference information is (as yet) available.

[6] Throughout the text, selection predictor effect sizes are denoted as d and refer to the standardized difference between the mean predictor score in the minority applicant and the majority applicant population. A negative d value therefore indicates a higher mean predictor score in the majority population as compared to the corresponding mean in the minority population.

no match for analytic-based methods when the latter methods can be applied.

Recognizing the limitation of EMOO methods, Song and Tang (2020) opted for a third angle for dealing with multiple minority groups. To this end, they used the procedure of De Corte et al. (2007) with some minor modifications to derive the PO selection systems and the corresponding PO front. This procedure implements the NBI approach of Das and Dennis (1998), using a sequential quadratic programming algorithm to solve the implied optimization problems.[7] Song and Tang presented an example application involving the situation with a single quality (i.e., the validity of the selection predictor) and two diversity objectives related to the Black and Hispanic AIR, respectively. Although their proposal has the advantage of using analytic methods to obtain a discrete approximation of the PO front, it is limited to the design of PO systems in single stage selection situations. Moreover, the proposal uses the NBI method that is known to often provide a poor and incomplete representation of the Pareto front in case of more than two objectives (see Burachik et al., 2017; Messac & Mattson, 2004; see also our comparison with the NBI approach below).

In summary, although previous research offers some steps toward the design of multiobjective PO selection systems, we still lack a reliable and generally applicable procedure for obtaining such systems. The next section describes a procedure that overcomes the limitations of these earlier efforts to obtain a discrete approximation of the PO front and the corresponding set of PO systems. The two-dimensional subspace procedures for generating informative subsets of the front in two-dimensions are presented in the subsequent section.

Approximating the Full Front of Multiobjective PO Selection Systems

Below we provide an elementary description of the procedure for obtaining PO selection systems for situations involving more than two objectives, independently of the particular combination of (multiple) quality and (multiple) diversity objectives. We refer to Appendix A for a detailed description of the procedure. To illustrate the multiobjective PO procedure, we intertwine the presentation with an example application and compare it to previous approaches to tackle this key issue.

Description of the Hybrid Multiobjective PO Approach

Similar to existing methods for deriving PO biobjective selection systems (e.g., De Corte et al., 2011; Rupp et al., 2020), the present procedure requires certain data as well as the delineation of the set of feasible selection systems. The latter set comprises all selection designs that are judged acceptable by the selection designer. It is defined by an appropriate series of constraints that typically express restrictions on the available predictors and the cost and timing of the selection as well as restrictions on acceptable predictor weights, eventual staging preferences, and so on. The required data relate to the proportional representation of the different applicant groups in the total candidate population, the intercorrelation, validity and effect size values (i.e., subgroup d for each minority group) of the predictors, the final selection rate and, in case of multistage selection, the range of acceptable between-stage retention rates. These data permit computing the value of the selection objectives for each feasible selection system using the formulas presented in De Corte et al. (2006). Note that the data requirements and boundary conditions for applying the new procedure are similar to those that govern the application of the biobjective PO approach. The only difference is that the new method requires effect size data for the predictors/criterion for each minority group, instead of for only one.

To obtain the PO trade-off front and the corresponding PO selection designs, we developed a new procedure that applies a hybrid multiobjective PO approach consisting of two stages. The first stage applies the SBG method of Mueller-Gritschneder et al. (2009), resulting in a first discrete representation of the PO front. Next, a novel method, henceforth referred to as the E-NBI method, that combines features from the ENNC and NBI methods is used to obtain a second discrete representation. A Pareto filter (a procedure that checks whether thus far obtained trade-offs are PO; cf. Messac & Mattson, 2004) is subsequently applied to the two solution sets to retrieve the final PO front and the corresponding PO selection designs.

Our new method combines the SBG and E-NBI procedures because the two procedures are quite complementary. The SBG method is particularly suited for obtaining a quite even representation of the boundary and the neighborhood of the boundary of the PO front, albeit without a guaranteed coverage of the entire inside part of the front (cf. Mueller-Gritschneder et al., 2009, p. 920), whereas the novel E-NBI method compensates this deficiency by resulting in a quite evenly populated discrete representation of the inner part of the PO front. When combined, the two methods promise a high-quality discrete representation of the full PO front such that the selection designer has a full overview of the acceptable selection designs when facing the decision on the preferred design. This is a key advantage as further shown in the section "Added Value of New Multiobjective PO Hybrid Approach over Previous Methods."

Apart from generating a high quality, discrete representation of the PO front in multiobjective selection design, the new hybrid method can also be extended to determine for each PO design the *corresponding convex formulation* of the design if the latter formulation exists. Based on a result presented in Das and Dennis (1998), De Corte et al. (2007) already introduced the notion of the corresponding convex formulation of a PO design in the context of the biobjective selection design situation. In this situation, the Pareto front is typically convex (i.e., the front curves outwardly) implying that a biobjective PO systems can also be obtained by solving a corresponding single objective optimization problem. The objective function of the latter problem is a convex combination (i.e., a linear combination of the objectives where the weights of the objectives in the linear combination are nonnegative and sum to one) of the two selection objectives, and the convex combination of the objectives is referred to as the corresponding convex formulation. As the result of Das and Dennis generalizes to multiobjective PO selection systems with a trade-off that lies in a convex part of the PO front, the present hybrid method extends the procedure of De Corte et al. to obtain the corresponding convex formulation of biobjective PO systems to the case of multiobjective PO designs in two steps. The method first

[7] Analytic methods to solve the MOOP lead to a series of single objective optimization problems that are typically solved by a nonlinear programming method, with sequential quadratic programming (cf. Fletcher, 2010) being the method of choice.

checks whether a corresponding convex formulation is possible and then determines the convex weights of the objectives in the formulation. As further discussed in the section "Sample-to Sample Cross-Validity of Multiobjective PO systems," this new development is of key importance when studying the shrinkage of the trade-off of multiobjective PO systems under cross-validation. In particular, we show that the weight of an objectives in the corresponding convex formulation of a PO design correlates substantially with the amount of shrinkage of the objective trade-off value.

Computer Program for the Hybrid Multiobjective PO Approach

We wrote a computer program that implements the new hybrid procedure for the design of PO selection systems involving up to five objectives. The output of the program contains a tabular inventory of the PO trade-offs and the corresponding PO selection designs. We also coded a simple R script that enables the visualization of the entire PO front in case of three objectives. In case of more than three objectives, the script can also be used to visualize the different three-dimensional parts of the full dimensional PO front.

Transparency and Openness

We provide a copy of the executable program that computes a discrete approximation of the full PO front and the corresponding PO selection designs for selection situations involving up to five selection objectives at http://users.ugent.be/~wdecorte/software.html. Instructions for using the program, several example input and output files as well as a copy of the R script to visualize the PO front are also available from the website. The program also computes the subsets of PO selection designs when implementing the two-dimensional subspace procedures detailed in the section Obtaining a Two-Dimensional Approximation of Part of the PO Front. All input values used in the following example applications as well as in the later reported simulation study are detailed in the Tables 1 and 2 and in the body of the article and we throughout adhered to the *Journal of Applied Psychology* methodological checklist. As the article does not report empirical research, it was not preregistered.

First Example Application

Data and Selection Situation Specification

Table 1 details the required data for the first example application aiming at the computation of the PO front for a three-objective selection with the predictor composite validity gauging the quality objective and the Black and Hispanic AIR expressing the two diversity objectives. The data values in the table are borrowed from Song and Tang (2020) and are based on previous meta-analytic studies, whereas the values for the proportional representation of the three applicant groups reflect the corresponding population statistics in the United States.

To be able to compare the results of the new procedure with the results reported in Song and Tang, the example assumes that only single stage selection designs with a .15 selection rate using a selection predictor composite where all available predictors receive a nonnegative weight are judged acceptable. For the same reason of comparability, the validity of the selection predictor composite is used to gauge the selection quality objective.

Table 1

Predictor/Criterion Population Data for the Three-Objective Selection Design Situation (Cf. Song & Tang, 2020)

Variable	d_B	d_H	1	2	3	4	5	6
Predictor								
1. Biodata	−.39	−.17	—					
2. Cognitive ability	−.72	−.79	.37	—				
3. Conscientiousness	.09	.08	.51	.03	—			
4. Structured interview	−.39	−.04	.16	.31	.13	—		
5. Integrity test	−.04	.14	.25	.02	.34	−.02	—	
Criterion								
6. Performance			.32	.52	.22	.48	.18	—

Note. d_B = standardized Black–White mean difference; d_H = standardized Hispanic-White mean difference. The proportional representation of the majority and the Black and Hispanic minority groups is .705, .170, and .125.

Results

Figure 1 illustrates the result of the new hybrid multiobjective PO approach when applied to the example data from Table 1. The figure depicts the full PO front together with the trade-off achieved by a fixed weight (FW) system assigning unit weight to each of the five predictors (cf. the triangle shaped hollow point) and the trade-off associated with a corresponding PO system (cf. the triangle shaped filled point). The FW system results in a validity of .57, with a substantial adverse impact against both Blacks (AIR Black = .45) and Hispanics (AIR Hispanic = .67). The corresponding PO system has a similar validity of .58 but a slightly higher AIR for Blacks (.51) and a substantially better AIR for Hispanics (.85). Summarized in compact notation, the FW System has a trade-off equal to (.57, .45, .67) and this trade-off is dominated by the trade-off of the corresponding PO system equal to (58, .51, .85) as each trade-off

Table 2

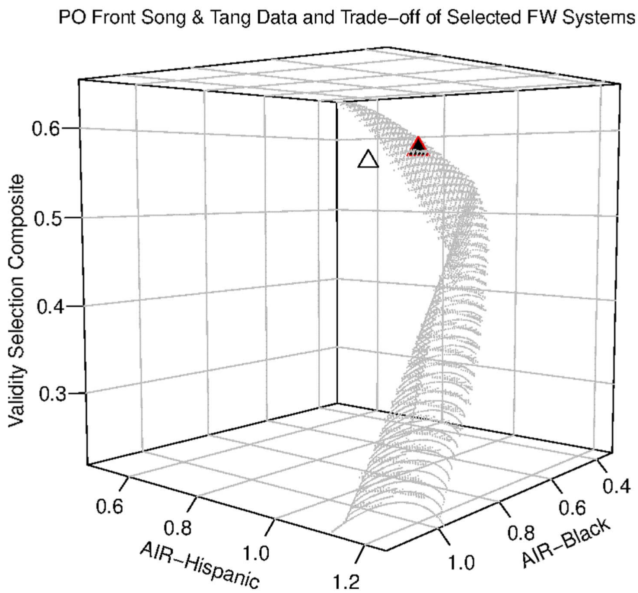
Predictor/Criterion Effect Sizes, Correlations, and Validities From the Army Data Set

Variable	d_B	d_H	d_F	1	2	3	4	5	6
Predictor									
1. Adjustment	.08	−.01	−.31	—					
2. Dependability	.23	.21	.31	.34	—				
3. Surgency	.05	.16	−.06	.57	.11	—			
4. g-factor	−.83	−.72	.06	.45	.10	.16	—		
Criterion									
5. Leadership	−.20	.10	−.35	.13	.12	.58	.36	—	
6. General soldiering	−.54	−.38	−.11	.20	.09	.28	.34	.03	—

Note. d_B = standardized Black–White mean difference; d_H = standardized Hispanic-White mean difference; d_F = standardized female–male mean difference; g-factor = general mental ability. Proportional representation of the majority, the Black, the Hispanic, and the Female minority groups is .60, .22, .03, and .15, respectively.

Figure 1

Pareto Front for the Three-Objective Selection Situation Corresponding to the Song and Tang (2020) Data



Note. The grey surface depicts the representative discrete approximation of the PO front obtained by the new hybrid method to compute multiobjective PO systems. The hollow triangle point shows the trade-off of the FW system assigning equal weight to the predictors, whereas the filled triangle point represents a corresponding PO trade-off that dominates the trade-off of the FW system. AIR = adverse impact ratio; PO = Pareto-optimal; FW = fixed weight. See the online article for the color version of this figure.

component of the FW system has a smaller value than the corresponding trade-off component of the PO system.

Besides a pictorial representation of the PO front, the new procedure also provides an overview of the corresponding PO selection systems. Figure 2 illustrates this by plotting the weights, with which the available predictors are combined to the selection composite predictor, across the full range of values for the quality (i.e., validity) objective. Note that the pattern of the predictor weights across the quality range is quite irregular, as may be expected in case of multiobjective selection design when the relation between the validity and the effect size of the predictors differs across the minority groups.

Added Value of the New Multiobjective PO Hybrid Approach Over Previous Methods

Comparison With NBI Approach

When discussing previous research, and the proposal of Song and Tang (2020) in particular, we noted that biobjective PO approaches in using the NBI method may often lead to a poor and even mistaken representation of the PO front when applied to selection situations involving more than two objectives. The results depicted in Figure 3 illustrate and substantiate this claim. The upper row panels of the figure show the PO front for the example selection scenario, detailed in the "Method" section above, as obtained when using the NBI (left upper row panel) or the present hybrid method (right upper panel). Note that the PO front obtained with the NBI method matches the front reported in Song and Tang (2020) who used the NBI method. Yet, comparing the fronts in the upper row panels of Figure 3 demonstrates that the fronts are quite different, with a substantial part of the front obtained by the NBI method being dominated by the PO front of the new method. The grey part of the NBI front shown in the lower left panel of Figure 3 details the dominated part of the NBI front, indicating that the largest part of the front derived by the NBI method is not PO.

Comparison With Repeated Application of the Biobjective PO Approach

Although NBI based methods often fail to derive a high-quality discrete representation of the PO front in case of more than two objectives, the repeated usage of the method, each time solving for the PO front of a different combination of two of the pursued objectives, may on aggregation eventually result in a more adequate representation of the full PO front. To investigate whether such a repeated usage may result in an adequate representation of the PO front in case of more than two-valued objectives, we again studied the Song and Tang example selection scenario and used the biobjective PO approach to first derive two sets of selection designs that each are PO with respect to the quality (i.e., the validity of the selection composite, see the section "Method" above) and only one of the two diversity objectives. The first set is PO with respect to the quality and the Black minority group AIR objectives; whereas the second set is PO with respect to the quality and the Hispanic AIR objectives. Next, we used the PO predictor weighting systems corresponding to the systems that are PO with respect to validity and Black (Hispanic) AIR to calculate the corresponding Hispanic (Black) AIR value, obtaining two sets of trade-offs with each trade-off having three components: the validity, the Black AIR and the Hispanic AIR trade-off value.

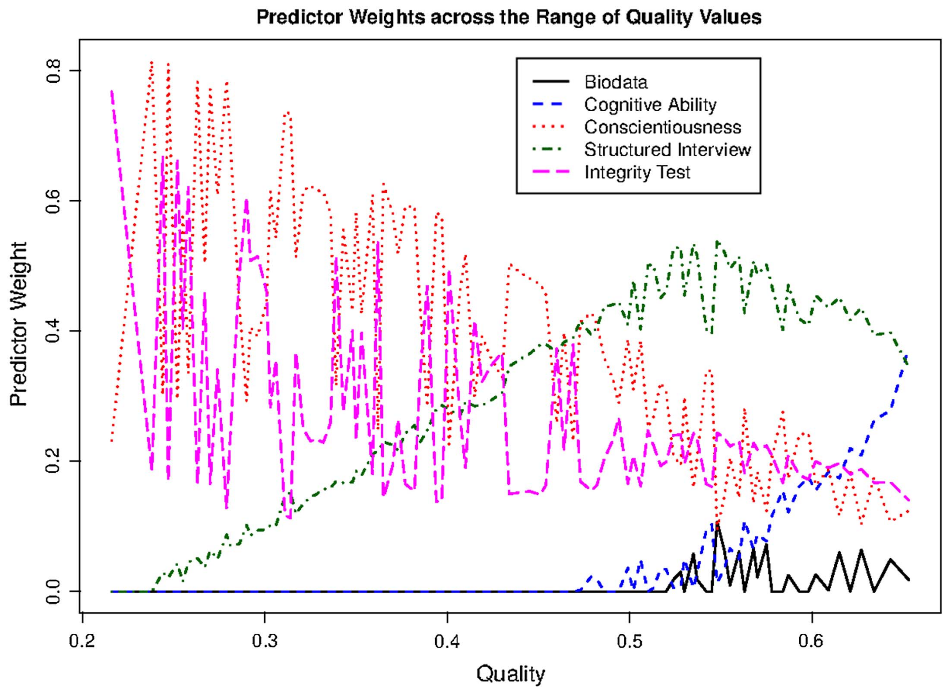
The bottom right panel of Figure 3 depicts the resulting two sets of trade-offs relative to the full PO front as computed by the new hybrid method. The red filled square points on the left side of the full PO front correspond to the predictor weighting systems that are PO with respect to the quality and only the Black AIR objective, whereas the blue filled circle point on the right side correspond to the predictor weighting systems that are PO with respect to the quality and only the Hispanic AIR objective. The results in the panel show that the repeated application of the biobjective PO approach in situations with more than two objectives uncovers only a (very) small and totally unrepresentative subset of the entire PO front (i.e., only part of the border of the PO front).

Conclusions and Discussion

The above disappointing results using existing methods are not specific to the studied example, but generalize to virtually all multiple (i.e., three or more) objective selection situations. In case of more than two objectives, dedicated methods such as the new hybrid multiobjective PO procedure are required to obtain a high-quality representation of the full PO front and the corresponding PO selection system designs. Approximating the PO front through repeated application of the biobjective PO approach is no alternative because the resulting selection designs are at best highly nonrepresentative. In fact, and this is a result of the example

Figure 2

Predictor Weights Corresponding to the Three-Objective PO Selection Systems (Song & Tang, 2020, Data)



Note. The PO systems are ordered from left to right in increasing quality trade-off value. PO = Pareto-optimal. See the online article for the color version of this figure.

application that needs additional qualification, it may very well be that the thus obtained systems are not even PO with respect to the full set of pursued goals (see Appendix C).

The comparison in the lower right panel of Figure 3 also verifies that focusing simultaneously on several minority groups instead of only one group at the time need not affect the selection rates and the adverse impact ratios that can be achieved by the PO selection designs. In the example, the PO trade-offs and, hence, the corresponding Black and Hispanic AIR values resulting from the two-objective PO approach are entirely part of the PO front obtained by solving for the systems that are PO with respect to both minority groups.

Similar to the two-objectives procedures, the present hybrid approach results in a discrete representation of the entire PO front and simple methods, such as filtering the list of obtained PO systems, can be used to select subsets that are of particular interest to the selection designer. Thus, in the above example application, the list of PO solutions can be reduced to (a) maintain only solutions with validity and/or AIR Black and AIR Hispanic values in a specific range (e.g., AIR values between .8 and 1.2), (b) to select the solutions that dominate the FW system assigning equal weights to the selection predictors, or (c) to focus on PO systems with trade-off component values (i.e., values for the individual elements of the trade-off such as the values for the quality or the AIR of a selection design) that meet or better certain preset values.

The above example also compares the outcome (i.e., the trade-off) of the FW system to that of the selection designs obtained by the present hybrid method (see Figure 1). The results indicate that the FW systems is typically dominated by several of the computed PO designs (i.e., has quality and diversity values that are all lower than the corresponding values of the PO designs) and, hence, is not PO, whereas the new hybrid multiobjective PO method retains only PO solutions. Our example application did not compare the PO systems to the regression-based predictor weighting scheme because the latter scheme is one solution of the derived PO systems. This will always be the case when the quality objective equals the validity of the selection predictor composite. If the quality objective is operationalized as the expected criterion performance of the selected applicants, as is necessary in case of multistage selections, the regression weighted selection system need not be PO because in that case the quality objective is not only a function of the validity of the selection composite predictor but also depends on the composition of the total applicant group and the effect sizes of the predictors and the criterion.

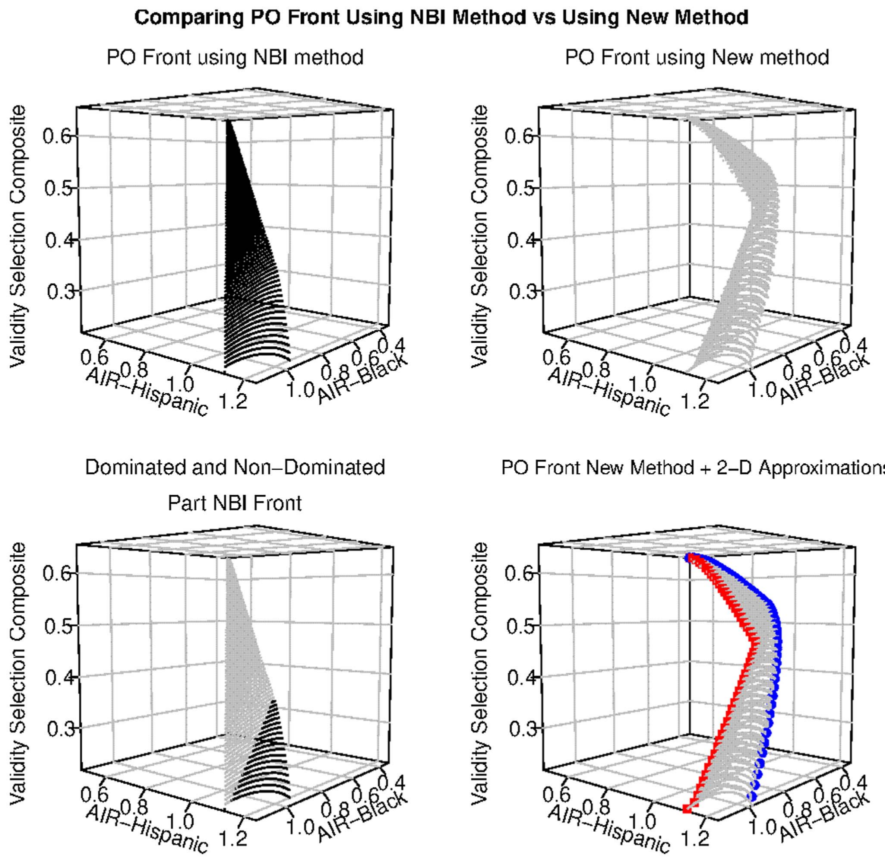
Obtaining a Two-Dimensional Approximation of Part of the PO Front

Rationale

Although a high-quality retrieval of the PO selection systems is vital in situations where both quality and diversity objectives are valued, the first example application illustrates that the retrieval inevitably leads to a further major issue that needs to be addressed: the decision on which of the identified PO systems to retain. Even in the example situation involving only three objectives and where the set of PO systems can still be visualized, the latter decision is not an

Figure 3

Three-Objective PO Front Obtained by the NBI and the New Method and Two Corresponding Two-Objective Solutions



Note. The upper row panels show the PO trade-off surface obtained by the NBI and the new hybrid method for the three-objective situation of the first example application. The black surface part in the bottom left panel shows the part of the NBI solution that is PO. The red square and blue circle points in the right bottom panel represent the PO trade-off curve of two different two-objective solutions (i.e., the validity and AIR Black and the validity and Hispanic AIR solutions, respectively). AIR = adverse impact ratio; NBI = normal boundary intersection; PO = Pareto-optimal. See the online article for the color version of this figure.

easy one. To assist the selection designer in the task, the next section presents three additional procedures that may guide the decision as to which PO selection design to retain.

Description of Two-Dimensional Subspace Procedures

This section introduces three two-dimensional subspace procedures for obtaining and visualizing certain particularly interesting subsets of the full PO front in two-dimensions. We again opt for an elementary description of the procedures and refer to Appendix B for details. We intertwine the description with two example applications, henceforth referred to as the example applications two and three.

The two-dimensional subspace procedures each compute a PO front related to a different pair of aggregate objectives. In all three computations, the first aggregate objective corresponds to a weighted average of the quality objectives,[8] whereas the second aggregate objective equals the average AIR or the average selection rate across the minority groups for the procedures one and two. In the third procedure, the second aggregate diversity objective equals the minimum AIR across the protected groups. In essence, the first two procedures are identical to the biobjective PO approach applied to an aggregate quality and diversity objective instead of to a single quality and a single diversity objective; whereas the third procedure requires the novel computational scheme detailed in Appendix B.

Although the first two procedures focus on an aggregate quality and an aggregate diversity objective, the resulting PO selection designs are also automatically PO with respect to each of the individual quality and diversity objectives and, hence, form a subset of the full PO front as obtained by the new hybrid method. To see this, consider a selection design, D, that is PO with respect to, for

[8] It is understood that the different quality objectives are all expressed in standardized units such that weighted addition makes sense. It is equally possible to consider the minimum quality across the (standardized) quality objectives as the aggregate quality objective.

example, the average of the individual quality and the average of the minority AIR objectives. In that case, no other design can dominate design D (i.e., can have values for all but one of the individual objectives that are equal to the corresponding values of design D and do better on the remaining objective). If such a system would exist it would necessarily have either a higher aggregate (i.e., average) quality or a higher aggregate diversity trade-off value (or both) than the corresponding trade-off values of the system D, which contradicts the fact that the system D is PO with respect to the aggregate objectives.

In contrast, the selection systems obtained by the third procedure, with the minimal AIR value across the minority groups as the aggregated diversity objective, are not also automatically PO with respect to the individual objectives. The initial computation of the designs that are PO with respect to the aggregate objectives is therefore followed by the additional step detailed in Appendix B to ensure that the finally retained systems are also PO with respect to each of the individual objectives.

Second Example Application

Data and Selection System Specification

We use a second example application to show the relationship in results between the hybrid procedure for deriving the full PO front and the two-dimensional subspace procedures This second example relies on a small subset of data from the U.S. Army Project A, previously used by Saad and Sackett (2002) and Sackett et al. (2003) in examinations of predictive bias. Project A examined a wide variety of predictors and criteria across 13 Army occupational specialties (i.e., jobs), with a total *N* of 5,044. More information about this database can be found in Campbell (1990) and Young et al. (1990). To illustrate our procedure, we use four predictors (detailed in Table 2) and two criteria (General Soldiering Proficiency and Effort and Leadership)[9] from the Project A data. This example application two illustrates the two-dimensional subspace procedures in the context of a three-objective, two-stage selection design situation with standardized expected performance for General Soldiering Proficiency as the single quality objective and the Black and Female AIR as the two diversity objectives.

As all PO approaches, the two-dimensional subspace procedures require the specification of the set of feasible selection systems. Example application two takes this set as comprising all two-stage selection systems (for more details on multistage selection system design, see De Corte et al., 2011) with a .15 selection rate, a retention rate between .30 and .60 after the first stage, and using only nonnegatively weighted predictor composites of the predictors one and four and the predictors two and three in the stages one and two, respectively

Results

The upper left, the upper right, and the lower left panels of Figure 4 show the PO front as obtained by applying the NBI, the SBG, and the novel E-NBI method, respectively, whereas the light (grey) dots in the lower right panel depict the discrete representation of the front generated by the new hybrid method that combines the results of the SBG and the E-NBI methods. The upper left panel confirms that the NBI method often results in a rather poor retrieval of the full extent of the entire PO front and the upper right panel illustrates the potential of the SBG method to generate a high-quality discrete representation of both the border and a large vicinity of the border of the PO front. However, the method does not always result in a full coverage of the entire PO front, as witnessed by the absence of PO trade-off points in the center part of the front. In turn, the lower left panel shows that the new E-NBI method compensates for this deficiency by offering a quite even retrieval of the inner part of the PO front, but the method is less good than the SBG method in providing a precise image of the border of the PO front.

The lower right panel of Figure 4 depicts the results of two of the two-dimensional subspace procedures on top of the full PO front[10] obtained by the hybrid procedure. The (red) "+" points show the trade-offs that are PO with respect to the quality and the minimum AIR objective, whereas the (blue) "×" points represent the trade-offs that are PO with respect to the quality and the average AIR objectives. As noted above, both sets of points are also PO with respect to all three individual objectives and are therefore part of the three objective PO front uncovered by the new hybrid method. The (blue) square shaped point in the panel exemplifies one of these trade-offs with values .41 and .94 for quality and average AIR, and values of .67 and 1.21 for the AIR of the first (Black) and the second (female) protected group, respectively. In turn, the (red) bullet PO trade-off has values .40 and .80 for the quality and minimum AIR, and values .80 and 1.01 for the AIR of the first (i.e., Black) and the second (i.e., female) protected group, respectively.

With a minimum AIR value across the minority groups of .80, the selection design corresponding to the latter trade-off obeys the 4/5ths rule for the AIR of all minority groups and may therefore be a valuable option for the selection designer when choosing between the different PO selection designs. Also, although both designs result in fairly equal values for the quality objective (i.e., .41 and .40 for the blue square and the red bullet trade-offs, respectively), the range between the AIR values is with a value of only .21 for the second PO trade-off considerable smaller as compared to the corresponding value of .54 for the first PO trade-off.

Third Example Application

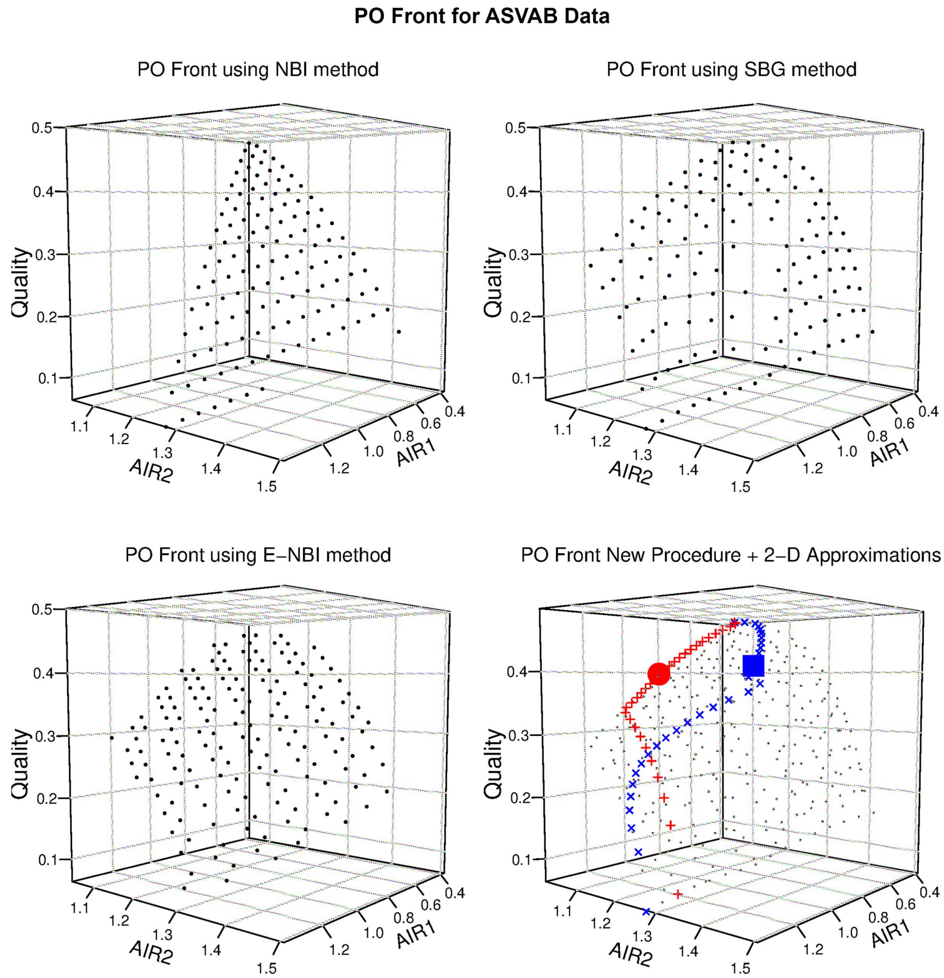
Purpose

Whereas our first two example applications mainly focus on the use of our new multiple objective PO approach in the context of multiple minority groups, we use example application three to illustrate the application of the subspace methods in a five-objective design situation. In this situation, the goal is to optimize two quality objectives (i.e., average performance of the selected applicants on General Soldiering Proficiency and Effort and Leadership) and three diversity objectives as represented by the AIR of the Black, Female, and Hispanic minority groups. In this

[9] As the data in Table 2 relate to incumbent instead of applicant groups, the values in the table are range restricted. Data to correct for this were not available, however.

[10] The lower panel of Figure 4 does not display the PO front for the quality and average selection rate objectives because the front is virtually identical to the PO front of the quality and average AIR objectives.

Figure 4
PO Trade-Off Front Obtained for the Second Example Situation Using ASVAB Data



Note. From left to right the top panels depict the PO front obtained using the NBI and the SBG method, whereas the left bottom panel represents the PO front obtained by the new E-NBI method. The right bottom panel shows the PO solutions of two of the two-dimensional subspace procedures superimposed on the PO front of the new hybrid method (cf. the grey surface). The red + and the blue × points correspond to the subspace procedure using the minimum AIR and the average AIR as the aggregate diversity objective, respectively. AIR = adverse impact ratio; NBI = normal boundary intersection; SBG = successive boundary generation; PO = Pareto-optimal; ASVAB = Armed Services Vocational Aptitude Battery. See the online article for the color version of this figure.

third example application, we also further explore the recommended use of our three subspace procedures given specific interests of the selection designer.

Data and Selection System Specification

Our third example application relies on the same subset of Project A data as our second example application. The selection system specification is also the same as in our second example: a .15 selection rate, a retention rate between .30 and .60 after the first stage, and using only nonnegatively weighted predictor composites of the predictors one and four and the predictors two and three in the stages one and two, respectively.

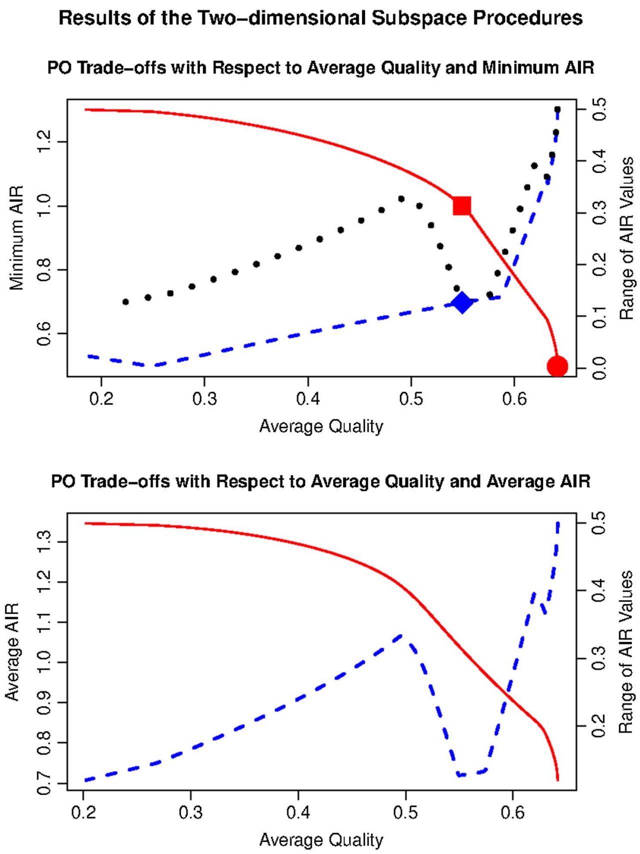
Results

Figure 5 summarizes the results obtained by two of the subspace procedures when applied to the above described third example situation involving a total of five objectives.[11] In the top panel, the solid (red) line shows the PO front for the average quality and the minimum AIR aggregate objectives, whereas the solid (red) line in the lower panel displays the PO trade-offs for the average quality

[11] We again omit the results of the subspace procedure, where the aggregate diversity objective is captured by the average selection rate because these results agree to a very high degree with the results from the procedure where average AIR gauges the aggregate diversity objective.

Figure 5

PO Trade-Off Curves for Two of the Subspace Procedures Showing the Relationship Between the Values for the Minimum AIR and the Average AIR Aggregate Diversity Objective



Note. The red line in the upper (lower) panel represents the PO trade-off curve of the subspace procedure using average quality and minimum (average) AIR for the objectives. The blue dashed line in the panels depicts the relationship between average quality and the range of the AIR values as obtained by the two subspace procedures. The blue dashed line of the lower panel is inserted as the dotted black line in the upper panel. The filled red square and filled red circle in the upper panel correspond to the PO trade-off with a value of one for the minimum AIR objective and a maximum possible value for the average quality objective, respectively. AIR = adverse impact ratio; PO = Pareto-optimal. See the online article for the color version of this figure.

and average AIR aggregate goals. Note that both fronts span a range of values for the two aggregate objectives thereby offering the possibility for the selection designer to focus on systems that show aggregate trade-off component values that meet eventual design preferences. As an example, consider the situation where the designer aims for a system resulting in the highest possible expected average quality and where all minority applicants are expected to have at least the same chance of being selected as the majority group candidates. In that case, the designer wants a system resulting in the best average quality subject to the requirement that the minimum AIR across the three minority groups is at least equal to one and the preferred system can immediately be identified as the (red) square point on the (red) line in the upper panel of Figure 5. The results in Appendix S4 of the online material show that the point corresponds to the PO selection system with aggregate (i.e., average) quality of .55 (and values of .77 and .33 for the leadership and general soldiering quality objectives, respectively) and minimum AIR of 1. (with values 1, 1.13 and 1. for the AIR of the Black, the Hispanic, and the Female protected groups, respectively).

The results in Appendix S4 of the online material further indicate that the preferred selection design first top-down selects the 60% candidates with the highest predictor composite score in the first stage, assigning weights of .51 and .49 to the Adjustment and *g*-factor (general mental ability) predictors, and finally top-down selects one quarter of the retained candidates based on their second stage composite score obtained by assigning weights of .23 and .77 to the second stage Dependability and Surgency predictors. As the design is expected to result in hiring candidates from the minority groups at least as frequently as the majority group candidates while losing no more than .09 standard units on the aggregate quality dimension as compared to the aggregate quality maximizing design (cf. the design marked by the red circle trade-off in the upper panel, with aggregate quality trade-off value of .64), it might be quite appealing to choose the design among all other PO designs.

Both panels also display the relationship between the aggregate quality and the range of the AIR values in the PO trade-offs, defined as the difference between the largest and the smallest value of the AIR trade-off components. Thus, the dashed blue line in the upper panel represents the relationship when aggregate quality and minimum AIR are the two aggregate objectives, whereas the corresponding line in the lower panel depicts the relationship when the aggregate objectives are the average quality and the average AIR. We added these lines to inspect which of the two operationalizations of the aggregate diversity objective (i.e., aggregate diversity operationalized as the average vs. the minimum AIR across the groups) is likely to result in a more homogeneous treatment of the different minority groups as reflected by the range in the AIR values. For reasons of comparison, the relationship between the aggregate quality and the range of AIR values for the systems that are PO when the average AIR represents the aggregate diversity objective is also depicted in the upper panel (cf. the dotted black line in the upper panel). Note that the latter line either touches or lies above the dashed blue line across the entire aggregate quality dimension thereby confirming the earlier obtained finding that the individual AIR values of the different minority group are typically more alike when the minimum AIR expresses the aggregate diversity as compared to when the average AIR is used to represent the objective.

Discussion

We offer three two-dimensional subspace procedures, each focused on a different aggregate diversity objective (e.g., average AIR vs. minimum AIR across groups) to assist selection system designers when evaluating the merits of different high dimensional PO systems. As illustrated in the example application three, the subspace procedures provide the selection designer with a manageable grasp of the possible PO selection designs in case that the number of valued objectives exceeds three.

Under which conditions should selection system designers use these procedures? We advise using the subspace procedures in combination with the new hybrid multiobjective PO procedure when the number of objectives remains modest. In that case, the results of the subspace procedures can, in principle, also be obtained from the output of the hybrid procedure, albeit considerably less systematically and in most cases also much less detailed. Note that the subspace procedures are no substitute for the recovery of the full dimensional PO front. So, their stand-alone usage is reserved only for occasions where the recovery of the full front poses excessive computational challenges or is expected to result in an all to overwhelming list of PO selection designs.

One might also wonder why we present these three two-dimensional subspace procedures. We acknowledge that several other two-dimensional subspace procedures are possible, (e.g., focusing on the minimum quality across several quality objectives). We retained the present three because current practice seems to be more often confronted with selection situations involving multiple minority groups than with situations with only one minority group and several different quality objectives. Also, as compared to other possible subspace procedures, the procedures focusing on the average AIR and the average selection rate across the minority groups have the advantage that the resulting designs are automatically PO with respect to the individual diversity objectives. Although the third subspace procedure (i.e., the minimum AIR across the minority groups to represent the aggregate diversity objective) requires additional steps to guarantee designs that are also PO for all individual objectives, the procedure has another important advantage. As shown in the example applications two and three, the minimum AIR subspace procedure leads to PO designs that, for comparable (aggregate) quality values, treat the different minority applicant groups more equally than the other two subspace procedures. So, if homogeneity in the AIR and, hence, also in the selection rates of the different groups is an issue, then this is the subspace procedure of choice.

Sample-to-Sample Cross-Validity of Multiobjective PO Systems

Rationale

As noted in the introduction section, the biobjective PO, the new hybrid multiobjective PO, and the subspace procedures result in designs that are PO, and therefore are better than any other feasible design, in terms of the *expected* trade-off of the valued objectives. However, this might change under cross-validation. When both PO and alternative feasible designs are applied in future selections involving a finite sized applicant pool it is possible that the validation trade-off (i.e., the trade-off achieved in the new, validation applicant pool) of the PO systems no longer outperforms the validation trade-off of (some of) the alternative designs. This is even more likely when (a) the initial computation of the PO systems is based on calibration sample rather than on population values of the predictor/criterion effect sizes and correlations and (b) the alternative designs do not depend on the idiosyncrasies of the calibration data (as is the case for FW selection designs). Although previous cross-validation studies (e.g., De Corte et al., 2022; Song et al., 2017) reported encouraging results on the cross-validity of PO as compared to FW designs, the results are limited to biobjective selection situations and may therefore not generalize to the multiobjective PO designs as obtained by the new hybrid method. To address these issues, we investigate the sample-to-sample cross-validity of the multiobjective PO systems via a small-scale simulation study.

The present study focuses on the sample-to-sample cross-validity of multiobjective PO systems because this is the type of cross-validity that is of particular interest in applied settings (Cattin, 1980; De Corte et al., 2022). More specifically, the study assesses the merits of multiobjective PO systems, derived on the basis of calibration applicant sample (summary) data, when applied in a new validation applicant sample. As shown by De Corte et al. (2022), a full-blown study of the sample-to-sample cross-validity of PO and alternative selection designs requires a dedicated methodology and considerable computational resources and is therefore out of scope for the present study. Instead, we aim to answer the key question as to whether the validation trade-off of seven alternative selection designs is still dominated by the validation trade-off of the PO designs that dominate the alternative selection designs in the calibration condition. The seven alternative selection designs are the FW selection design assigning equal weights to the selection predictors, and six expert weighing (EW) designs. To obtain the latter designs, we contacted twelve Industrial-Organizational practitioners in our network who specialize in selection, asking them to propose a set of weights for operational use in a setting where both validity and diversity were valued, given the *population* validities, intercorrelations, and effect sizes of the predictors in the different minority groups. Seven of these are Society for Industrial and Organizational Psychology fellows; two are Society for Industrial and Organizational Psychology past-presidents. All twelve responded. As some offered virtually identical weightings, we identified six meaningfully different expert weights for inclusion in the cross-validity study.

The simulation study also presents results on the shrinkage under cross-validation of the PO, the FW and the EW system trade-off values. Similar to previous studies, shrinkage refers to the difference between the average (across replications) calibration trade-off and the average validation trade-off of a systems. Although a detailed analysis of the shrinkage issue falls outside the scope of this article, we also study the relationship between the degree of shrinkage in the trade-off components (each corresponding to a selection objective) and the weighting of the objectives in the corresponding convex formulation of the PO system (cf. the section "Description of the Hybrid Multiobjective PO Approach"). Uncovering this relationship is important because it sheds light on the shrinkage pattern of the trade-off components one may expect to obtain under cross-validation.

Simulation Method

We use simulation methods to study the cross-validation issue with respect to the first example selection situation where five predictors are available to select candidates, with a selection ratio of .15, from an applicant group comprising both Black and Hispanic candidates with a proportional representation of .170 and .125, respectively. In the example, the single quality goal is represented by the validity of the selection composite predictor, whereas the two diversity objectives are gauged by the Black and Hispanic AIR, respectively.

Each simulation consists of three steps that are repeated 5,000 times for four levels of the size of the calibration and applicant pool: 100, 200, 500, and 1,000. In the first step, we generate two

Table 3

Proportion of Validation PO Trade-Offs That (a) Dominate, (b) Are Dominated, and (c) Are Incomparable to the Validation Trade-Off of the Corresponding Alternative Design System for Different Calibration/Validation Sample Sizes

Alternative system	PO trade-off dominates				PO trade-off is dominated				Incomparable trade-offs			
	100	200	500	1,000	100	200	500	1,000	100	200	500	1,000
FW	.24	.27	.34	.40	.21	.11	.04	.01	.55	.62	.63	.59
EW1	.23	.22	.23	.25	.25	.18	.09	.05	.52	.60	.68	.70
EW2	.18	.18	.18	.20	.32	.23	.13	.08	.50	.60	.69	.72
EW3	.22	.24	.31	.38	.23	.13	.05	.02	.55	.62	.64	.60
EW4	.13	.15	.16	.19	.37	.30	.16	.11	.49	.56	.68	.69
EW5	.18	.21	.30	.37	.27	.16	.05	.02	.56	.63	.65	.61
EW6	.28	.29	.36	.43	.18	.10	.03	.01	.55	.61	.61	.56

Note. The expert weight systems are detailed in Appendix D. The proportions in, for example, the row EW3 and the columns with sample size 200 indicate that, with calibration and validation sample sizes of 200, the proportion that the PO systems that dominate the EW3 system in the calibration condition continue to dominate in 24% (across the 5,000) repetitions) of the validation samples, are dominated in 13% of the validation samples, and are incomparable in 62% of the validation samples. PO = Pareto-optimal; FW = fixed weight; EW = expert weight.

random samples of predictors/criterion score data from the mixture normal distribution with population mean and correlation structure detailed in Table 1. The first of these two samples is the calibration sample, whereas the second is the validation sample. In the second step, the calibration predictors/criterion correlations and effect sizes are computed from the calibration sample data and subsequently used as input to the new hybrid multiobjective PO method to determine a discrete representative set of 231 selection systems that are PO with respect to the expected trade-off of the three valued goals. In computing the PO systems, it is understood that only predictor composites assigning nonnegative weights to the predictors are admissible. In the third step, the PO system designs (i.e., the weights with which the predictors are combined to the PO selection composite) are applied to both the calibration and validation sample data, resulting in the calibration and validation trade-off of each of the PO systems. The step also computes the calibration and validation trade-off of the FW and the six EW selection designs.

To study the key question whether PO systems that dominate the FW and EW systems in the calibration condition continue to do so in the validation condition, we first tabulated for each of the seven alternative designs the PO systems that dominate the system in the calibration condition. Next, we counted for each alternative system the number of corresponding dominating PO systems with a validation trade-off that (a) dominates, (b) is dominated by and (c) is incomparable to the validation trade-off of the alternative system. Note that the trade-off of two systems is *incomparable* when some of the trade-off values of the first system are higher than the corresponding trade-off values of the second system, whereas the reverse is the case for some other trade-off values.[12]

To evaluate the shrinkage under sample-to-sample cross-validation, we calculate the difference of the average (across the 5,000 replications) calibration and validation trade-off component values (i.e., the quality component as represented by the average criterion performance of the selected applicants and the AIR Black and the Hispanic AIR diversity components as represented by the average of the Black and the Hispanic AIR) of the 231 calibration PO systems and compare these 231 PO shrinkages to the corresponding shrinkages for the alternative systems. In addition, we report the standard deviation (across the repetitions) of the validation trade-off components of both the PO and the alternative systems. Finally, we examine the relationship between the shrinkage in the trade-off components of the PO systems and the average weight of these components (i.e., the selection objectives) in the corresponding convex formulation of the systems (cf. the section “Description of the Hybrid Multiobjective PO Approach”). For each PO system, the latter averages are computed across the repetitions where the PO system has a corresponding convex formulation. Based on previous results obtained for the biobjective selection situation (De Corte et al., 2022 and Song et al., 2017), we expect that the shrinkage in a trade-off component of a PO system is related to the weight of the corresponding objective in the corresponding convex formulation.

Results of Simulations

Table 3 reports the results on whether calibration-based PO systems that dominate an alternative selection system also outperform the alternative system under cross-validation. To this end, the table presents the proportion of comparisons (averaged across the 5,000 repetitions) between cross-validated alternative and corresponding cross-validated PO systems where the validation PO trade-off (a) still dominates, (b) is dominated, and (c) produces indeterminate findings (i.e., neither the PO nor the alternative system dominates such that the trade-off of the PO and the alternative system are incomparable).

The first key result is that the PO systems retain an advantage under cross-validation over the corresponding alternative systems for calibration/validation sample sizes of 200 or more. From this sample size onwards, the odds that the validation trade-off of the PO systems dominates versus is dominated by the validation trade-off of the corresponding alternative designs grows from 27/11 (sample size 200) to 34/4 (sample size 500) and 40/1 (sample size 1,000) for

[12] As an example, consider two systems, with trade-off value for the validity, AIR Black and AIR Hispanic objectives of (.45, .96, 1.02) for the first system and trade-off equal to (.50, .1.05, .90) for the second system. In that case, the two systems have an incomparable trade-off as the validity and the AIR Black trade-off value of the second system are higher than the corresponding trade-off value of the first system, and the AIR Hispanic trade-off value of the first system is higher than the corresponding trade-off of the second system.

the FW system, and from 22/18 to 26/9 and 30/5 on average for the EW systems. The results also vary somewhat across the EW designs, with one of the designs (i.e., EW4) performing at least equally well as the corresponding PO designs for calibration/validation sample sizes of up to 500.

As a second key result, the proportions in Table 3 show that the validation trade-off of the alternative systems and the corresponding PO systems are incomparable in more than half of the cases. Also, the proportion that the trade-offs are incomparable remains quite stable across the different sample sizes of the calibration/validation condition.

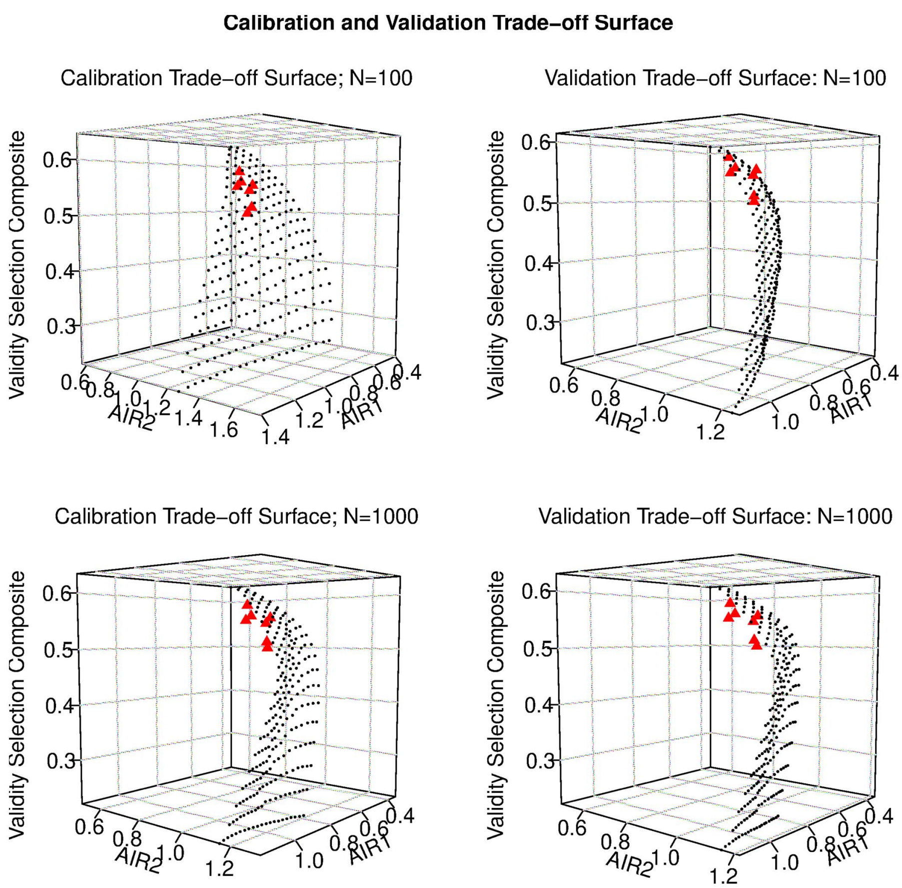
Appendix S5 of the online material provides full numerical results of the shrinkage analysis, whereas Figure 6 shows the average trade-off of the seven alternative (red triangle points) and the 231 calibration-based PO systems (black dots) when applied to the calibration and the validation sample, respectively. The upper row panels in Figure 6 correspond to the smallest calibration/validation applicant pool size of 100, whereas the bottom panels refer to the largest sample size condition of 1,000. The results reported in Appendix S5 of the online material further show that the average validation trade-off of the alternative selection designs EW3, EW5, and EW6 are dominated by the average validation trade-off of one or more PO systems from the smallest calibration/validation sample size of 100 onwards, whereas the same is true for the designs EW1 and EW2 for a sample size of at least 200.

Figure 7 details the results on the relationship between the shrinkage in the trade-off components and the weight of the components in the corresponding convex formulation. The rows of the figure refer to the three trade-off components, the columns correspond to the four studied sample sizes and each panel depicts the relationship between the shrinkage and the average convex formulation weight for the particular combination of sample size and trade-off component. The correlation between shrinkage and average convex formulation weight is added at the top of each panel.

Together, the results confirm the expectation that shrinkage in the PO trade-off components is inversely related to the calibration/

Figure 6

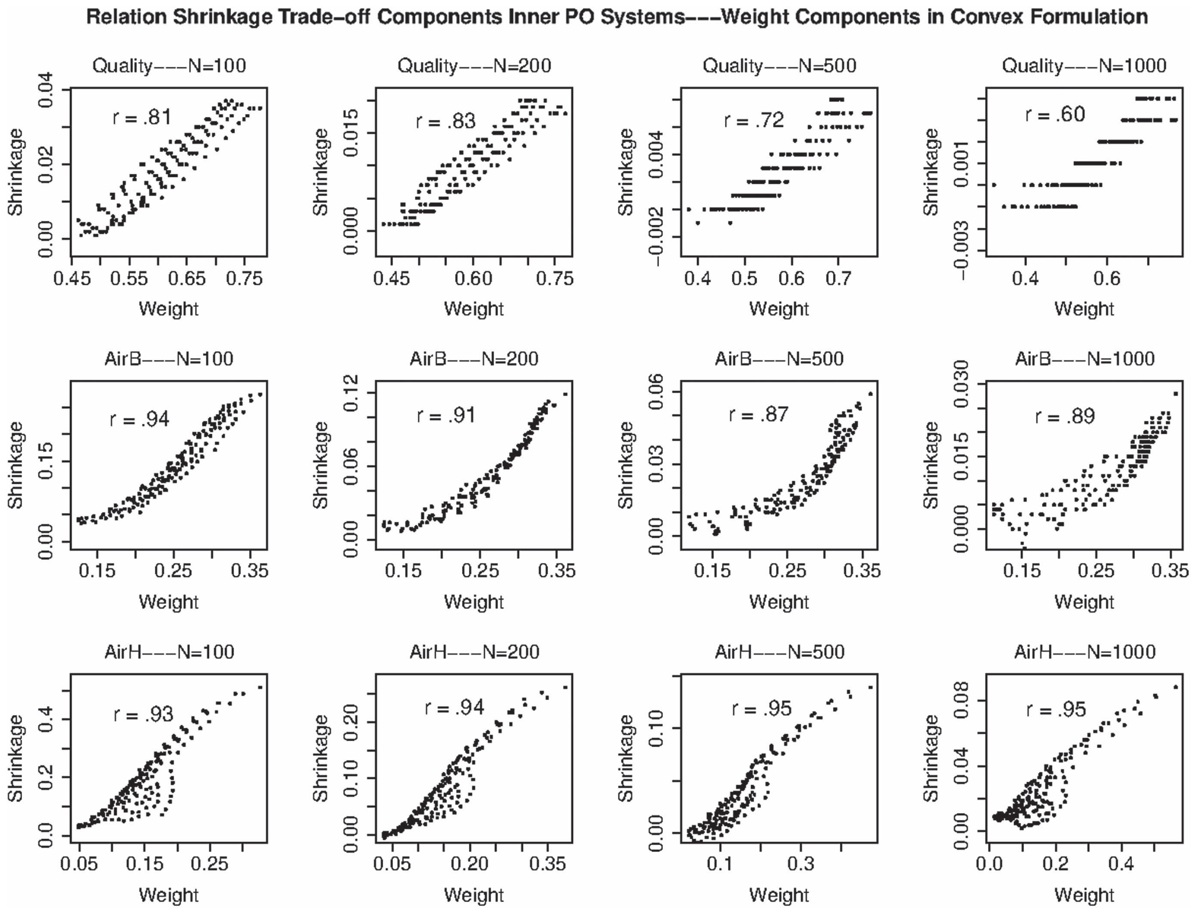
Calibration and Validation Trade-Off Surface for Sample Size Conditions of 100 and 1,000



Note. The small black dots represent the average (across 5,000 replications) trade-off of the calibration-based PO system when applied to the calibration sample (left panels) and the validation sample (right panels). The large red squares show the average trade-off of a FW and six EW systems (see text) in the calibration and validation sample. Calibration/validation sample size equal to 100 and 1,000 for the upper and the bottom panels respectively. AIR = adverse impact ratio; PO = Pareto-optimal; FW = fixed weight; EW = expert weight. See the online article for the color version of this figure.

Figure 7

Relation Between the Shrinkage in PO Trade-Off Components and the Weight of the Component in the Corresponding Convex Formulation



Note. The small (black) points refer to the PO systems. PO = Pareto-optimal; AIR = adverse impact ratio.

validation sample size and that the shrinkage in the components of the PO systems is proportional to the weight of the component in the corresponding convex formulation. Table 4 provides further insight in the relation between shrinkage and sample size by detailing the average (across systems) shrinkage in the trade-off components for the four sample size conditions for each of the three types of studied selection designs: the PO, the FW, and the EW designs. For each combination of trade-off component, sample size and selection design type, the table also lists the range (across systems) of the standard deviation (across repetitions) of the validation trade-off component values. The values for the average shrinkage of the trade-off components listed in the table show that only PO systems show meaningful shrinkage, but the shrinkage diminishes substantially for larger sample size conditions. Also, the amount of shrinkage is substantially lower for the quality trade-off component as compared to the AIR B and AIR H trade-off components. Whereas shrinkage is limited to PO systems, the values for the range of the standard deviation of the validation trade-off components show that both PO and alternative selection designs result in variable trade-offs across repetitions, especially in case of the AIR B and AIR H diversity trade-off components. However, the variability is more substantial and more diverse for the PO designs as compared to the alternative designs because it reflects not only the variability in the calibration condition across the repetitions as is the case for all designs, but also because the different PO systems vary in the weight of the trade-off components in the corresponding convex formulation.

Discussion

For reasons of feasibility, our cross-validation study focused only on two issues: the shrinkage in the trade-off of PO and alternative selection designs under cross-validation, and the likelihood that calibration-based PO systems that dominate an alternative system continue to do so under cross-validation. The latter issue is no doubt the key one because it speaks to the real practical value of multiobjective PO selection system design. As witnessed from the beginning, the presently used procedure to settle the issue, although appropriate, is far from sensitive as contrasting the validation trade-off of different multiobjective selection systems is very often impossible because the trade-offs are incomparable. To resolve the issue, the much more demanding approach discussed in De Corte et al. (2022) should be adopted. Yet, despite the low sensitivity of

Table 4
Average Shrinkage Trade-Off Components and Range of the Standard Deviations of Validation Trade-Off Components of PO and Alternative Selection Systems for Varying Applicant Pool Sizes

	Average shrinkage			Standard deviation		
Sample size	Quality	AIR B	AIR H	Quality	AIR B	AIR H
			PO selection systems			
100	0.02	0.14	0.19	0.06–0.15	0.39–0.86	0.55–1.04
200	0.01	0.05	0.09	0.04–0.13	0.25–0.54	0.35–0.66
500	0.00	0.02	0.04	0.03–0.11	0.15–0.33	0.22–0.39
1,000	0.00	0.01	0.03	0.02–0.09	0.11–0.24	0.15–0.27
			FW selection system			
100	0.00	0.02	0.01	0.07	0.45	0.62
200	0.00	0.00	0.00	0.05	0.30	0.41
500	0.00	0.00	−0.01	0.03	0.18	0.25
1,000	0.00	0.00	0.00	0.02	0.12	0.18
			EW selection systems			
100	0.00	0.01	0.01	0.07–0.07	0.41–0.55	0.61–0.76
200	0.00	0.00	0.00	0.05–0.05	0.28–0.36	0.39–0.50
500	0.00	0.00	−0.01	0.03–0.03	0.16–0.22	0.25–0.31
1,000	0.00	0.00	0.00	0.02–0.02	0.12–0.16	0.17–0.22

Note. Quality refers to the average job performance of the selected candidates, whereas AIR B and AIR H correspond to the AIR trade-off value for the Black and the Hispanic minority group applicants. Average shrinkage refers to the average (across the systems) of the average (across the repetitions) shrinkage in the trade-off components. Standard deviation ranges refer to the range (across the systems) of the standard deviation (across the repetitions) of the validation trade-off component values. Shrinkage refers to the average (across the repetitions) shrinkage in the trade-off components. Single standard deviations refer to the standard deviation (across the repetitions) of the validation trade-off component values. AIR = adverse impact ratio; PO = Pareto-optimal; FW = fixed weight; EW = expert weight.

the present procedure, we obtained results that favor PO selection system designs over the FW and three of the six EW designs for sample size conditions of 200 and more. The result is of paramount importance to organizational practices when developing PO systems in that it implies using sufficiently large calibration sample size conditions to obtain the systems. The presently obtained sample size of 200 may be an overestimate, however, because the comparison between the PO and the EW disadvantages the PO systems, in that the EW systems are based on the population values of the predictor/criterion correlations and effect sizes, whereas the PO systems rely only on variable calibration sample predictor/criterion data. However, asking experts for a design proposal based on the calibration correlation and effect size data in each of the 5,000 different repetitions is impossible so that the present findings only offer a lower bound approximation on how PO systems perform under cross-validation as compared to the EW systems.

We also studied the shrinkage issue to conform with some of the previous related studies that, in following the vast body of research on the cross-validity of regression-based prediction, focus primarily on this feature (e.g., Song et al., 2017). As expected, FW and EW systems show virtually no shrinkage, whereas PO systems are prone to considerable shrinkage especially when the systems derive from small sized (i.e., sample size of less than 200) calibration samples. As the size of the calibration sample increases, shrinkage shrinks, eventually attaining, for calibration sample sizes of at least 500, approximately the same low level as obtained for the FW and EW systems.

Apart from studying shrinkage as related to the size of the calibration sample size, the present study introduces a novelty in that it is the first to systematically investigate the relationship between shrinkage in the PO trade-off components and the weight of the components in the corresponding convex formulation. The latter weights express the exact importance of the components in the system instead of the crude, ordinal approximation of the importance used in the previous studies (e.g., Song et al., 2017). However, the improved shrinkage analysis does not invalidate, but rather confirms the main results of these former studies that the shrinkage of a trade-off component is proportional to the weight of the component in the corresponding convex formulation of the PO system. As noted above, this is a valuable result for the organizational practice of selection design because it addresses the expected shrinkage pattern of the trade-off components without the need to conduct cross-validity analyses. Finally, despite the shrinkage, additional analyses show that the cross-validated FW and five of the six EW system are dominated by at least one cross-validated, shrunken PO system.

The above results illustrate that shrinkage analyses provide an ambiguous answer to the key cross-validity issue on whether calibration PO systems still outperform otherwise constructed designs under cross-validation because the computation of PO systems, similar to regression-based prediction systems, capitalizes on the idiosyncrasies of the calibration data and therefore necessarily results in systems showing higher shrinkage than systems, such as the FW and EW systems, that do not depend on these idiosyncrasies. When using the shrinkage criterion (i.e., the amount of shrinkage), PO systems, just as regression-based systems will always be the loser. Therefore, our present interest for the shrinkage is, apart from reasons of conformity, not motivated by the potential of shrinkage results to settle the key cross-validation issue, but rather by the fact that the shrunken trade-offs (cf. the trade-offs represented in the right panels of Figure 6) convey more accurate information than the initial calibration trade-offs on the actual trade-offs one may expect to obtain in future applications of the systems.

General Discussion

In response to key concerns in selection research and practice to extend the PO approach to settings with multiple minority groups the article proposes a new hybrid multiobjective PO approach for obtaining a representative discrete collection of the selection designs that are PO with respect to both several quality and multiple diversity objectives. As a second contribution, we present three two-dimensional subspace procedures that result in designs that are PO with respect to all individual objectives as well as with respect to the aggregated quality objective and an aggregate diversity objective. As a consequence, the resulting designs are part of the design collection obtained by the hybrid multiobjective PO approach. Finally, we introduced a considerably improved procedure for studying the relationship between the shrinkage under cross-validation of the PO trade-off components and the exact weight of the components in the corresponding convex formulation of the PO system.

Although selection situations involving multiple minority groups and only a single quality objective are at present of primary concern, we stress that all our new procedures can also deal with situations involving both multiple quality objectives and multiple minority groups as well as with situations where multiple quality goals are valued but only a single minority group is present. The methods to address these different situations are identical and the knowledge that these procedures can be applied in these different situations may thus also enable selection designers to select candidates with respect to different quality objectives such as average job performance, peak job performance, organizational citizenship behavior, etc.

We used several example applications to illustrate the procedures, emphasizing that they should be used in conjunction whenever the number of objectives is not too high and that the stand-alone usage of the subspace procedures is reserved for situations where the hybrid multiobjective PO method cannot be applied. If possible, the new hybrid method should always be used because it is the only procedure that provides a complete and representative overview of the entire set of systems that result in an expected quality and diversity trade-off that is PO for all the individual objectives. It is therefore the only procedure that informs the designer on the complete gamut of eligible PO selection designs. We also showed how the subspace procedures may assist the selection designer when choosing between the different PO designs. As an example, the third example application illustrated that the solution systems obtained by the minimum AIR subspace procedure may be of particular interest in situations where the designer aims for selection systems with homogeneous selection rates for the different minority groups. We also presented a simple procedure to narrow the results of the hybrid and the subspace procedures to a (considerably) smaller collection of PO systems that may be of special interest to the practitioner.

We further demonstrated the added value of the new hybrid multiobjective PO approach over the existing methods for PO selection design when both are applied to a multiobjective selection situation involving a single quality and multiple diversity goals. We showed that repeated usage of the biobjective PO approach, each time focusing on the quality goal and one of the diversity objectives, at best results in only a small subset of the selection designs uncovered by the hybrid multiobjective PO method. Yet, some of the systems obtained by the biobjective approach may show a quality and a diversity trade-off for the single minority group that is better than any of the corresponding two-objective trade-offs obtained by the hybrid multiobjective PO method. This may raise the concern that systems that are PO for multiple minority groups could somehow be harmful for one or the other minority group. The concern is ill-founded, however, because the occurrence of these better biobjective trade-offs is an artifact due to the fact that the objectives in the biobjective approach are operationalized with respect to only the majority and a single minority group instead of with respect to the full set of groups. When applied to the multiobjective situation these systems will either belong to the hybrid method solution set or be dominated by some of these solutions. If the candidate pool of a selection comprises applicants from multiple minority groups and the designer values the diversity objective with respect to each of the minority groups, it follows by definition that none of the systems derived by the biobjective PO approach can offer a better expected quality and diversity trade-off for all objectives than the PO systems obtained by the hybrid method.

As to the key question whether the multiobjective PO systems retain an advantage over alternative FW and EW systems when applied in future applicant samples, our cross-validity results suggest that this is indeed the case when the PO systems are based on calibration samples of 200 and more. The result is not particularly clear cut, however. Although PO systems, that dominate an alternative system in the calibration condition, continue to do so considerably more frequently in the validation condition as compared to the reverse (i.e., by a factor ranging between 1.2 and 40, depending on the type of alternative system and the calibration sample size), it is also found that the validation trade-off of the alternative and the corresponding PO systems are very often incomparable, precluding a decision on which type of system is best.

Two caveats are in order. First, neither the hybrid nor the subspace procedures resolve the final choice problem as to the preferred selection design. Identical to the biobjective PO approach, the final choice of the preferred PO system remains the discretion of the selection designer because the choice requires a valuation of the different objectives that may vary across application contexts and/or organizations. That is, it is a matter of the values of the organization whether one is willing to accept a given reduction in quality (e.g., 1%, 5% or 10%) for a given increase in diversity. If the selection designer has certain preferences or relevant information to guide the valuation process a multiattribute decision making (MADM) method (see, e.g., Rao & Lakshmi, 2021, for an overview of the MADM methods) can be used to obtain the preferred selection design. In the absence of such information, methods based on the concept of the "knee" of the Pareto frontier (Das, 1999) and an extension of the related measure of the distance between the PO trade-offs and the hyperplane containing the individual optima of the different quality and diversity objectives (cf. Petchrompo et al., 2022) may offer an alternative. Yet, whatever the chosen selection system, the present hybrid and subspace procedures guarantee that the finally chosen design is PO and, hence, cannot be bettered by any other feasible selection system.

Second, designing PO systems for multiobjective selection situations requires advanced multiobjective optimization methods that are even more complex than the methods invoked in the biobjective PO approach. So, the introduction of the present procedures in mainstream practice poses a challenge. To bridge that gap we make available an executable program that performs both methods and R scripts to visualize the results. Some examples and directions on using the program are also available from http://users.ugent.be/~wdecorte/software.html.

Conclusion

The article presents methods and tools to assist selection designers in developing selection systems that are PO with respect to multiple selection objectives. The methods are particularly useful in situations where applicants come from multiple minority groups and the designer aims for a selection system that optimally balances the valued quality and the diversity objectives. In contrast to any other approach, the methods have, especially in conjunction, the benefit that they reveal the full gamut of possible PO designs the designer can choose from, and, at the same time point to certain subsets of PO systems that may be of particular interest in a given selection situation.

References

Andrews, J. S., & Geden, M. (2021, April). Algorithmic employee selection considering multiple protected classes using evolutionary many objective optimization. In Q. C. Song & S. Wee (Co-chairs), *Multi-objective optimization in the workplace 2.0* [Symposium]. Conducted at the 36th Annual Convention of the Society for Industrial and Organizational Psychology.

Burachik, R. S., Kaya, C. Y., & Rizvi, M. M. (2017). A new scalarization technique and new algorithms to generate Pareto fronts. *SIAM Journal on Optimization*, *27*(2), 1010–1034. https://doi.org/10.1137/16M1083967

Campbell, J. P. (1990). An overview of the army selection and classification project (project A). *Personnel Psychology*, *43*(2), 231–240. https://doi.org/10.1111/j.1744-6570.1990.tb01556.x

Cattin, P. (1980). Estimation of the predictive power of a regression model. *Journal of Applied Psychology*, *65*(4), 407–414. https://doi.org/10.1037/0021-9010.65.4.407

Das, I. (1999). On characterizing the knee of the pareto curve based on normal-boundary intersection. *Structural Optimization*, *18*(2–3), 107–115. https://doi.org/10.1007/BF01195985

Das, I., & Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, *8*(3), 631–657. https://doi.org/10.1137/S1052623496307510

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 181–197. https://doi.org/10.1109/4235.996017

De Corte, W., Lievens, F., & Sackett, P. R. (2006). Predicting adverse impact and mean criterion performance in multistage selection. *Journal of Applied Psychology*, *91*(3), 523–537. https://doi.org/10.1037/0021-9010.91.3.523

De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology*, *92*(5), 1380–1393. https://doi.org/10.1037/0021-9010.92.5.1380

De Corte, W., Lievens, F., & Sackett, P. R. (2022). A comprehensive examination of the cross-validity of pareto-optimal versus fixed-weight selection systems in the biobjective selection context. *Journal of Applied Psychology*, *107*(8), 1243–1260. https://doi.org/10.1037/apl0000927

De Corte, W., Sackett, P. R., & Lievens, F. (2011). Designing Pareto-optimal selection systems: Formalizing the decisions required for selection system development. *Journal of Applied Psychology*, *96*(5), 907–926. https://doi.org/10.1037/a0023298

De Corte, W., Sackett, P. R., & Lievens, F. (2020). Robustness, sensitivity, and sampling variability of Pareto-optimal selection system solutions to address the quality-diversity trade-off. *Organizational Research Methods*, *23*(3), 535–568. https://doi.org/10.1177/1094428118825301

Fletcher, R. (2010). The sequential quadratic programming method. In G. Di Pillo & F. Schoen (Eds.), *Nonlinear optimization* (pp. 165–214). Springer.

Khaledian, K., & Soleimani-Damaneh, M. (2015). A new approach to approximate the bounded Pareto front. *Mathematical Methods of Operations Research*, *82*(2), 211–228. https://doi.org/10.1007/s00186-015-0510-4

Messac, A., & Mattson, C. A. (2004). Normal constraint method with guarantee of even representation of complete Pareto frontier. *AIAA Journal*, *42*(10), 2101–2111. https://doi.org/10.2514/1.8977

Mueller-Gritschneder, D., Graeb, H., & Schlichtmann, U. (2009). A successive approach to compute the bounded Pareto front of practical multiobjective optimization problems. *SIAM Journal on Optimization*, *20*(2), 915–934. https://doi.org/10.1137/080729013

Nagy, M., Mansour, Y., & Abdelmohsen, S. (2020). Multi-objective optimization methods as a decision making strategy. *International Journal of Engineering Research & Technology*, *9*, 516–522. https://www.researchgate.net/publication/340362482_Multi-Objective_Optimization_Methods_as_a_Decision_Making_Strategy

Petchrompo, S., Coit, D. W., Brintrup, A., Wannakrairot, A., & Parliko, A. K. (2022). A review of pruning methods for multi-objective optimization. *Computers & Industrial Engineering*, *167*, Article 108022. https://doi.org/10.1016/j.cie.2022.108022

Peterson, Z., & Morris, S. B. (2021, April). Optimized for whom? Do Pareto-optimal solutions generalize across groups? In Q. C. Song, & S. Wee (Co-chairs), *Multi-objective optimization in the workplace 2.0* [Symposium]. Conducted at the 36th Annual Convention of the Society for Industrial and Organizational Psychology.

Rao, R. V., & Lakshmi, R. J. (2021). Ranking of Pareto-optimal solutions and selecting the best solution in multi- and many-objective optimization problems using R-methodology. *Soft Computing Letters*, *3*, Article 100015. https://doi.org/10.1016/j.socl.2021.100015

Roth, P. L., Van Iddekinge, C. H., DeOrtentiis, P. S., Hackney, K. J., Zhang, L., & Buster, M. A. (2017). Hispanic and Asian performance on selection procedures: A narrative and meta-analytic review of 12 common predictors. *Journal of Applied Psychology*, *102*(8), 1178–1202. https://doi.org/10.1037/apl0000195

Rupp, D. E., Song, Q. C., & Strah, N. (2020). Addressing the so-called validity-diversity trade-off. Exploring the practicalities and legal defensibility of Pareto-optimization for reducing adverse impact within personnel selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *13*(2), 246–271. https://doi.org/10.1017/iop.2020.19

Saad, S., & Sackett, P. R. (2002). Investigating differential prediction by gender in employment-oriented personality measures. *Journal of Applied Psychology*, *87*(4), 667–674. https://doi.org/10.1037/0021-9010.87.4.667

Sackett, P. R., Laczo, R. M., & Lippe, Z. P. (2003). Differential prediction and the use of multiple predictors: The omitted variables problem. *Journal of Applied Psychology*, *88*(6), 1046–1056. https://doi.org/10.1037/0021-9010.88.6.1046

Sayin, S. (2000). Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, *87*(3), 543–560. https://doi.org/10.1007/s101070050128

Schlüter, M., Egea, J. A., & Banga, R. (2009). Extended ant colony optimization for non-convex mixed integer nonlinear programming. *Computers & Operations Research*, *36*(7), 2217–2229. https://doi.org/10.1016/j.cor.2008.08.015

Shukla, P. K. (2007). On the normal boundary intersection method for generation of efficient front. In Y. Shi, G. D. van Albada, J. Dongarra, & P. M. A. Sloot (Eds.), *Computational science-ICCS 2007. Lecture notes in computer science* (pp. 310–317). Springer.

Song, Q. C., & Tang, C. (2020, April). Adverse impact reduction for multiple subgroups: A Pareto-optimization approach. In Q. C. Song & S. Wee (Co-chairs), *Multi-objective optimization in the workplace: Addressing adverse impact in selection* [Symposium]. Conducted at the 35th Annual

Convention of the Society for Industrial and Organizational Psychology, Austin, TX.
Song, Q. C., Wee, S., & Newman, D. A. (2017). Diversity shrinkage: Cross-validating pareto-optimal weights to enhance diversity via hiring practices. *Journal of Applied Psychology*, *102*(12), 1636–1657. https://doi.org/10.1037/apl0000240
Young, W. Y., Houston, J. S., Harris, J. H., Hoffman, R. G., & Wise, L. L. (1990). Large-scale predictor validation in Project A: Data collection procedures and data base preparation. *Personnel Psychology*, *43*(2), 301–312. https://doi.org/10.1111/j.1744-6570.1990.tb01560.x
Waller, N. G. (2008). Fungible weights in multiple regression. *Psychometrika*, *73*(4), 691–703. https://doi.org/10.1007/s11336-008-9066-z
Wee, S., Newman, D. A., & Joseph, D. L. (2014). More than *g*: Selection quality and adverse impact implications of considering second-stratum cognitive abilities. *Journal of Applied Psychology*, *99*(4), 547–563. https://doi.org/10.1037/a0035183

Appendix A

Hybrid Procedure for Deriving the Full PO Front and Corresponding PO Selection Designs

The hybrid procedure for computing the full PO front and the corresponding PO selection system designs applies two approaches for obtaining a discrete representation of the PO front: the SBG and the novel E-NBI approach. The resulting two approximations of the full PO front are combined in an overall discrete approximation and passed through a Pareto filter (cf. Messac & Mattson, 2004) to obtain the final solution front. The SBG approach has the advantage that it provides a high quality discrete representation of the PO front boundary, but it may fail to cover the entire extent of the Pareto front (cf. Mueller-Gritschneder et al., 2009, p. 920). To compensate for this, the hybrid procedure also applies the new E-NBI approach that combines features from the enhanced normalized normal constraint (ENNC) (Messac & Mattson, 2004) an the NBI (Das & Dennis, 1998) methods. Details on both approaches are discussed next. Here, we note that the hybrid procedure first transforms the initial multiobjective maximization problem $\max_{\mathbf{x}\in\mathcal{X}} \mathbf{g}(\mathbf{x})$ to the equivalent standard multiobjective minimization problem $\min_{\mathbf{x}\in\mathcal{X}} -\mathbf{g}(\mathbf{x}) = \min_{\mathbf{x}\in\mathcal{X}} \mathbf{f}(\mathbf{x})$, where $\mathbf{f}(\mathbf{x}) = (-g_1(\mathbf{x}), \ldots, -g_n(\mathbf{x}))'$ and $\{g_1, \ldots, g_n\} = S$ denotes the total set of n quality and diversity goals. Also, $\mathcal{X}$ represents the set of feasible values for the decision variables (also referred to as design variables or design parameters) $\mathbf{x}$ with the elements of $\mathbf{x}$ corresponding to the cutoff scores applied to the stage predictor composites and the weights of the predictors in the stage predictor composites. The restrictions that define the set $\mathcal{X}$ typically refer to bound constraints on (a) the predictor weights used in forming the stage-specific predictor composites (e.g., only nonnegative weights) and (b) the intermediate retention and the final selection rate of the selection. Additional constraints related to the staging of the predictors, the timing and cost of the selection procedure and so on may also be included in the set of restrictions that define $\mathbf{x}$.

SBG Approach

The SBG approach sequentially generates a discrete approximation of the complete PO front boundary in n steps. In the first step, the approach solves for the minima of the individual objectives $f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})$, resulting in the set $\{\mathbf{f}^{*1}, \ldots, \mathbf{f}^{*n}\}$ of individual minimum objective performances (IMOPs). The set comprises the discretized PO fronts, $\overline{\delta F}^{fj}$, with $j = 1, \ldots, n$, of the n single element combinations from the set S, and each element has a corresponding compromise weight vector of the objectives, with (1, 0, …, 0)' the compromise weight vector corresponding to $\mathbf{f}^{*1}$ and (0, 0, …, 1)' the corresponding compromise weight vector for $\mathbf{f}^{*n}$. The IMOPs are also called anchor points, with the i-th anchor point, $\mathbf{f}^{*i}$, equal to $(f_1(\mathbf{x}^{*i}), \ldots, f_n(\mathbf{x}^{*i}))' = (f_1^{*i}, \ldots, f_n^{*i})'$ with $\mathbf{x}^{*i}$ the solution of minimizing the objective function $f_i(\mathbf{x})$. The anchor points define the *utopia* point $\mathbf{f}^U = (f_i(\mathbf{x}^{*1}), \ldots, f_n(\mathbf{x}^{*n}))'$, whereas the *nadir point* $\mathbf{f}^N$ equals $(f_1^N, \ldots, f_n^N)'$, where $f_i^N = \max_{\mathbf{x}\in\mathcal{X}} f_i(\mathbf{x})$.

From the second step onwards, the computation of the discretized PO fronts associated with the two- up to the n-element combinations of the objectives is based on (a) the discretized PO fronts obtained in the previous steps, suitably ordered in a matrix denoted as $\boldsymbol{F}_B$, and (b) the corresponding matrix, $\mathbf{W}_B$, of compromise weight vectors, with columns $\mathbf{w}_{B,l}$ having elements $\mathrm{w}_{B,l,i}$ and $i = 1, \ldots, n$ (see section 7.3 in Mueller-Gritschneder for additional details on the matrices $\boldsymbol{F}_B$ and $\mathbf{W}_B$). In the i-th step, the discretized Pareto front of each i-element combination of the objectives is generated, using an appropriate additional set of compromise weight vectors, $\overline{\mathcal{W}}$, with the density parameter D determining the cardinality of the set. Each of the additional compromise vectors $\mathbf{w}_k$ of $\overline{\mathcal{W}}$ is first mapped to a corresponding base point $\mathbf{b}_k = \mathbf{F}_B\mathbf{a}_k$, with $\mathbf{a}_k$ the solution of the linear program *minimize* $\mathbf{d}_k^{'}\mathbf{a}_k$ subject to $\mathbf{W}_B\mathbf{a}_k = \mathbf{w}_k$ and $\mathbf{a}_k \geq \mathbf{0}$, where the element $d_{k,l} = \sum_{i=1}^{n} |w_{B,l,i} - w_{k,i}|$. Given the base points $\mathbf{b}_k$ for an i-element combination of the objectives and the target trajectory vector $\mathbf{v}$, with elements $v_i = \max_{j=1,\ldots,n} f_i^{*j} - f_i^{*i}$, the PO objective vectors $\mathbf{f}_k^*$ of the corresponding PO front are obtained by solving the following nonlinear programming problem: $\min_{t,\mathbf{x}} t$ subject to $\mathbf{0} \leq \mathbf{v}t + b_k - \mathbf{f}(\mathbf{x})$ and $\mathbf{x} \in \mathcal{X}$.

New E-NBI Approach

The new E-NBI approach proceeds in two stages. The first stage applies the enhanced normal constraint method proposed in Messac and Mattson (2004) to generate a set of evenly distributed points on the enlarged utopia plane section that contains the orthogonal projection of the hypercube $\mathcal{H}$ enclosing the entire feasible space $\mathcal{F}$ equal to the set $\{\mathbf{f}(\mathbf{x})\}$ with $\mathbf{x} \in \mathcal{X}$. The second stage adopts the NBI method to obtain the corresponding PO objective value vectors. The first stage starts with the above detailed computation of the IMOPs. The IMOPs delineate a section of the utopia plane which is defined as the hyper plane that contains the IMOPs. Next, a series of benign nonlinear programs is solved to determine the maximum and minimum value of the α^j parameters (denoted as α_l^j and α_u^j, with $j = 1, \ldots, n$, the number of objectives) that correspond to the enlarged utopia plane section that contains the orthogonal projection of the hypercube $\mathcal{H}$. The generic form of these programs is: $\min_{\mathbf{f},\boldsymbol{\alpha}} \alpha^j (\max_{\mathbf{f},\boldsymbol{\alpha}} \alpha^j)$ subject to (a) $\sum_{i=1}^{n} \alpha^i = 1$, (b) $\mathbf{f}^U \leq \mathbf{f} \leq \mathbf{f}^N$ (c) $-\infty \leq \alpha^i \leq \infty, \forall i \in \{1, \ldots, n\}$, and (d) $(\mathbf{f}^{*r} - \mathbf{f}^{*k})'(\mathbf{f} - \mathbf{p}) = 0$, $\forall k \in \{1, \ldots, n\}$, $k \neq r$; where $\mathbf{p} = \sum_{i=1}^{n} \alpha^i \mathbf{f}^{*i}$ represents a point on the enlarged utopia plane section, and $r \in \{1, \ldots, n\}$. By

(*Appendices continue*)

varying α^j from α_l^j to α_u^j by even increments (related to the intended density D of the discretized representation of the PO front), with $\sum_{i=1}^{n} \alpha^i = 1$, we obtain a set of evenly distributed points $\mathbf{p}_l$ (with associated weights $\boldsymbol{\alpha}_l = (\alpha_l^1, \ldots, \alpha_l^n)'$) on the enlarged utopia plane section. Some of these points cannot lead to a feasible solution, and a further benign nonlinear program is repeatedly solved (i.e., for each objective) for each generated $\mathbf{p}_l$ point to check that the point can result in a feasible nondominated objective vector solution. The generic form of the latter program is $\underset{\mathbf{f}}{min}\ C$ subject to (a) $\mathbf{f}^U \leq \mathbf{f} \leq \mathbf{f}^N$; (b) $(\mathbf{f}^{*j} - \mathbf{f}^{*k})'(\mathbf{f} - \mathbf{p}_l)$ for $\forall k \in \{1, \ldots, n\}$, $k \neq j$; and (c) $\sum_{i=1}^{n}(f_i - f_i^{*r}) - \sum_{i=1}^{n} |f_i - f_i^{*r}| - q$, where $r \in \{1, \ldots, n\}$, q is a small number and C is a constant.

Using the weights $\boldsymbol{\alpha}_l$ of the retained $\mathbf{p}_l$ points instead of the weights proposed by Das and Dennis (cf. the β weights in section 5.1 of Das and Dennis), the second stage of the E-NBI approach applies the NBI as well as the modified NBI method of Shukla (2007) to compute the corresponding set of PO objective value vectors. More specifically, the following nonlinear program is solved for each weight vector $\boldsymbol{\alpha}_l$: $\underset{t,\mathbf{x}}{max}\ t$ subject to (a) $\boldsymbol{x} \in \mathcal{X}$, and (b) $\boldsymbol{\Phi\alpha}_l - t\boldsymbol{\Phi}\mathbf{e} - \mathbf{f}(\boldsymbol{x}) + \mathbf{f}^U = \mathbf{0}$, where the i-th column of $\boldsymbol{\Phi}$ equals $\mathbf{f}^{*i} - \mathbf{f}^U$ in case of the NBI method, whereas the modified NBI solves the above problem with constraint (b) nonnegative. To avoid a piling up of the border front solutions only the solutions that are identical under both methods are retained.

Further Comments

Throughout the new hybrid procedure all nonlinear programs are solved using a sequential quadratic programming algorithm (Fletcher, 2010). To assure a manageable computational load, the program that implements the hybrid approach incorporates some built-in limitations with respect to the density of the intended discrete approximation of the PO front. However, the program also automatically applies the below discussed two-dimensional low space procedure. In all applications, the number of design variables must be sufficiently large (i.e., exceed the maximum number of binding constraints) to guarantee solution of the PO front. As a rule of thumb, the number of design parameters should not be less than the number of objectives plus the number of selection stages.

During the development of the procedure, four additional analytic and two heuristic methods for computing the PO front were also tested. The analytic methods were the full enhanced normalized normal constraint method (Messac & Mattson, 2004), the classic NBI method (Das & Dennis, 1998), the new scalarization technique of Burachik et al. (2017), and the proposal of Khaledian and Soleimani-Damaneh (2015). The heuristic methods were the evolutionary multiobjective (EMOO) method proposed by Deb et al. (2002) and the extended ant colony optimization algorithm presented by Schlüter et al. (2009). The tests revealed that none of the additional methods added significantly to the quality of the solutions obtained by the present hybrid approach. When judged by the quality standards for the discrete representation of PO fronts described by Sayin (2000), the heuristic methods perform rather poorly in terms of coverage and uniformity as quite a few parts of the PO front remain often unexplored, whereas others are quite densely but also unevenly populated. Except for the NBI and, albeit to a lesser degree, the relaxed NBI methods that both tend to result in a poor representation of the total PO front, the other three methods score better in terms of coverage and uniformity, but not to a degree that warrants inclusion of the methods in the present hybrid approach.

Appendix B

Two-Dimensional Subspace Procedure

The two-dimensional subspace procedures solve for three sets of selection systems that are PO with respect to an aggregated quality and an aggregated diversity objective as well as with respect to the individual quality and diversity objectives. The three sets share the same aggregated quality objective, equal to the average of the individual quality objectives, but differ with respect to the nature of the aggregated diversity objective equal to either the average AIR, the average selection rate or the minimum AIR across the minority groups.

The computation of the sets with the average AIR and the average selection rate as aggregate diversity objective proceeds in three stages: (a) the calculation of the individual maximum performance objectives $\mathbf{g}^{*\bar{q}} = (\bar{q}^+, \bar{a}_-)'$ and $\mathbf{g}^{*\bar{a}} = (\bar{q}_-, \bar{a}^+)'$ with $\bar{q}$ and $\bar{a}$ the aggregate quality and aggregate diversity objective, (b) the computation of the nonboundary PO selection systems, and (c) the application of a Pareto filter. The associated nonlinear programs for the first stage are $\underset{\mathbf{x}\in\mathcal{X}}{max}\ \bar{q}(\mathbf{x})$ and $\underset{\mathbf{x}\in\mathcal{X}}{max}\ \bar{a}(\mathbf{x})$, respectively, and $\underset{\mathbf{x}\in\mathcal{X}}{max}\ \bar{q}(\mathbf{x})$ subject to $\bar{a}(\mathbf{x}) = a$, with a a regularly spaced value in the interval $(\bar{a}_-, \bar{a}^+)$, for the second stage. The Pareto filter in the third stage eliminates eventual non-PO solutions.

The computation of the PO aggregate objectives $\bar{q}$ and $\underset{i=1,\ldots,l}{min}\ a_i$, with a_i the diversity for group I, also adopts the above detailed three steps, but the associated nonlinear programs are different. Also, a fourth step is required to check that the solutions obtained for the aggregate objectives are also PO with respect to all individual diversity and quality objectives. Thus, to obtain the maximum for the $\underset{i=1,\ldots,l}{min}\ a_i$ aggregate diversity objective the following nonlinear program is solved: $\underset{t,\mathbf{x}\in\mathcal{X}}{max}\ t$ subject to $t \leq a_i(\mathbf{x})$ for $\forall i = (1, \ldots, l)$. Using a_+ to denote the solution value of $\underset{i=1,\ldots,l}{min}\ a_i$ obtained in the first step and a_- for the value of $\underset{i=1,\ldots,l}{min}\ a_i$ associated with the solution of $\underset{\mathbf{x}\in\mathcal{X}}{max}\ \bar{q}(\mathbf{x})$, the programs solved in the second step have the form $\underset{\mathbf{x}\in\mathcal{X}}{max}\ \bar{q}(\mathbf{x})$ subject to $a_i(\mathbf{x}) \geq a$ for $\forall i = (1, \ldots, l)$ with a a regularly spaced value in the interval $(a_-, \ a_+)$. After applying the Pareto filter a final step checks and eventually updates the retained aggregate objective solutions, with associated individual objective values $(a_1^k, \ldots, a_l^k, q_1^k, \ldots, q_m^k)$ and minimum AIR value of a_-^k for the k-th solution, to ensure that they are also PO with respect to the individual quality and diversity objectives. To implement the step the following nonlinear program is solved repeatedly (i.e., for $i = 1, \ldots, l$) for each retained solution k: $\underset{\mathbf{x}\in\mathcal{X}}{\max} \sum j = \underset{j\neq i}{1}, \ldots, l\ (a_j(\mathbf{x}) - a_j^k)$, subject to (a) $a_i(\mathbf{x}) = a_-^k$, (b) $a_j(\mathbf{x}) \geq a_-^k$ for $\forall j = 1, \ldots, j$ and $j \neq i$, and (c) $q_j(\mathbf{x}) \geq q_j^k$ for $\forall j = 1, \ldots, m$. If any of the l programs of a solution k results in a significant positive value of at least .0001, the step is repeated starting from the solution design parameter values of the program.

(*Appendices continue*)

Appendix C

Fungible PO Systems

The design parameter values of a PO selection system solve a set of nonlinear and eventually linear (equality and inequality) equations. Using $\mathbf{t}$ to denote the vector of trade-off values of a PO system, $\mathbf{g}(\mathbf{x})$ for the functions that relate the design parameters $\mathbf{x}$ to the different pursued objectives, and $\mathbf{x}_t$ for the design parameter values of a PO solution, some of the nonlinear equations express that $\mathbf{g}(\mathbf{x}_t) = \mathbf{t}$, whereas still other nonlinear and linear (equality and inequality) equations may correspond to restrictions that define the set of feasible design systems. With an increasing number of design parameters (i.e., with an increasing number of available predictors in the selection stages), the system of equations may count many different solutions, implying that $\mathbf{x}_t$ is not uniquely defined.

In the regression context, the set of predictor weights leading to the same prediction validity is known as the set of fungible weights (e.g., Waller, 2008), and the title of the present appendix extends this usage of the term "fungible" to the context of PO selection system design by defining a fungible PO system as a system for which different sets of design parameter values result in the same trade-off value for the objectives and by referring to these sets of design parameter values as fungible design parameter value sets. Note that the issue of fungible PO systems is quite different from issues related to obtaining only locally optimal instead of globally optimal solutions when solving for the PO systems. Even globally optimal and, hence, truly PO designs may be fungible.

Given two selection situations, the first involving two-valued goals and the second characterized by the same two goals as well as a third goal, the occurrence of fungible design parameter value sets for systems that are PO in the first situation implies that not all different members from the set will also result in a system that is PO in the second situation. For some members, the value of the third goal, as implied by the design parameter values, may be less than the corresponding value of the third goal obtained when solving for the second situation using a three-objective optimization method to obtain the PO front.

Appendix D

Details of the EW Alternative Designs

The six EW systems reflect the alternative design proposed by six experts in the field when asked for a nonstatistical weighting of the predictors to balance the three valued selection objectives of example application one: the validity of the selection composite and the Black and Hispanic AIR. To decide on the weighting, the experts were given the population values of the predictor/criterion correlation and effect size values detailed in Table 1. The resulting weights for the predictors one to five are, for Expert 1, EW1: 1, 1, 5, 4, 2; for Expert 2, EW2: 1, 2, 4, 5, 4; for Expert 3, EW3: 3.5, 2.5, 3, 4, 2; EW4: 1, 1, 3, 5, 1; EW5: 2, 2, 2, 4, 0; and EW6: 5, 1, 3, 4,2.

■