Measurement issues in the evaluation of projects in a project portfolio

Alec Morton[a]

[a]*Department of Management Science, Strathclyde Business School, University of Strathclyde, Glasgow G1 1XQ, United Kingdom. E-mail: alec.morton@strath.ac.uk*

Abstract

A common problem arising in many domains is how value the benefits of projects in a project portfolio. Recently, there has been some attention given to a decision analysis practice whereby analysts define a value function on some criterion by setting 0 as the value of the worst project. In particular, Clemen and Smith have argued this practice is not sound as it gives different results from the case where projects are "priced out", and makes a strong implicit assumption about the value of not doing a project. In this paper we underscore the criticism of this way of using value functions by showing it can lead to a rank reversal. We provide a measurement theoretic account of the phenomenon, showing that the problem arises from using evaluating projects on an interval scale (such as a value scale) whereas to guard against such rank reversals, benefits must be measured on at least a ratio scale. Seen from this perspective, we discuss how the solution proposed by Clemen and Smith, of explicitly reformulating the underlying optimisation problem to allow for explicitly non-zero values, addresses the issue, and explore in what sense it may be open to similar problems. In closing we discuss what lessons from practice may be drawn from this analysis, focussing on settings where the Clemen and Smith proposal may not be the most natural way of modelling.

Keywords: Decision analysis, Portfolio Decision Analysis, multicriteria decision analysis (MCDA)

---

1. Introduction

All organisations have to decide which activities (out of a set of possible activities) they will initiate, which they will continue, and which they will stop.

Supporting such decisions by evaluating projects and assessing the benefits of possible portfolios as a function of the component projects represents a significant part of the professional work of many analysts. The application of decision analysis in such settings is often referred to as "Portfolio Decision Analysis": Portfolio Decision Analysis seeks to provide frameworks for performing portfolio analysis in a way which allows a rigorous treatment of issues of value and uncertainty. Kleinmuntz (2007) and Phillips and Bana e Costa (2007) have provided overviews of past work, and personal statements of what good practice looks like. The edited volume by Salo, Keisler and Morton (2011) provides a range of perspectives on theory and practice in this area.

A common practice in Portfolio Decision Analysis in a multicriteria setting is to define criterion-specific value functions by setting 0, the baseline for measurement, as the value of the worst item, as recommended by Phillips and Bana e Costa (2007, p 57) and implicitly recommended Kirkwood (1996); see for example Kirkwood's example of Section 8.1 and the discussion in footnote 3 of Clemen and Smith (2009). These scores are then summed up across criteria to give an overall project score, and summed across the constituent projects to give an overall portfolio score. Clemen and Smith (2009) (henceforth, "CS") argue that it is not a sound approach to analysis as it gives different results from the case where projects are "priced out", and makes a strong implicit assumption about the value of not doing a project. As a solution, they advocate explicitly assigning project-specific value scores for not doing particular projects.

A number of scholars have sought to build on the work of CS. In particular, Liesiö and Punkka (2014) have explored how to elicit a suitable baseline value and also provide techniques for exploring sensitivity of the solutions recommended by a particular decision model to uncertainty about the appropriate baseline values. de Almeida et al. (2014) explore three consequences of rescaling the valuations of projects in a portfolio setting, which they dub the portfolio size effect, the baseline effect, and the issue of consistency across different aggregation sequences; and de Almeida and Vetschera (2012) explore a similar issue in the context of the PROMTETHEE multicriteria method.

CS demonstrate by citing and describing in some detail numerous published examples that this practice is widespread. Notably, they cite an application to

2

capital budgeting in healthcare (Kleinmuntz and Kleinmuntz, 1999); an application to project selection in a telecommunications company (Linstedt et al., 2008); and a military portfolio optimisation (Parnell et al., 2002). They comment that that this list is not exhaustive and they "have found similar issues in many studies" (p 262). This author has had the same experience. For example, in the context of R&D prioritisation, Morcos (2008) recommends expressing scores "along a value scale with intervals from 0 (least desirable) to 100 (most desirable)" (p 77). (Examination of the figures in this paper suggests that he departs from this rule in the case of one criterion, "profitability" but adheres to it for the "risk" and "reliability" criteria.) Similarly Baltussen and Niessen (2006) present an illustrative example of multicriteria techniques to priority setting in healthcare (in which the task is to rank order projects with a view to deciding what subset of them to undertake) in which the zero has been set at the level of the least preferred project on three of four criteria (for the fourth criterion, "severity of disease", the projects are given between one and four stars, and the zero is set at one star). The reason for the pervasiveness of this practice seems to be that setting the zero at the level of the least preferred alternative is standard, unproblematic and uncontested advice in the context where one is choosing a single item from a set. Many authors generalise this advice to the setting where the task is to choose multiple objects from a set.

In this paper we will study the setting of baselines from a measurement theoretic point of view. Seen from this perspective, the issue is one which has surfaced in different forms in the literature over the years. In particular, Rao and Weiss (1986) provide a numerical exploration (in the context of a particular application) of the number of possible ways of ranking items in benefit/ cost order when the benefit scores of individual items are translated by a constant. Roberts (1990) provides an introduction to the theory of measurement and explores how it may be relevant in the context of combinatorial optimization problems: he observes that one can only be sure that the solution to a combinatorial optimization problem will be invariant with respect to monotonically increasing transformations of its objective function coefficients when the optimizing algorithm uses only ordinal information about these coefficients (for example, a greedy algorithm).

The contribution of this paper will be to study the practice discussed by CS

3

and answer the following questions.

1. Can we get a clear view of what goes wrong when we use value functions to value projects multicriteria portfolio decision analysis?
2. What are the specific characteristics of value functions which give rise to this problem? Might this problem apply more generally when project benefits are assessed in some other way? In what way is this a specifically *portfolio* problem?
3. Why does the solution proposed by CS "work" to resolve this problem? Is the solution proposed by CS vulnerable to other similar and related problems?
4. How important is the separability of the rule for combining project benefit values as overall portfolio benefits in this analysis? Do the results we obtain for this rule give insights into what should happen in a more general setting?
5. What, practically speaking, can we learn from this analysis about how to perform better and more robust analysis?

The structure of the paper will be as follows. Section 2 will answer question 1. Section 3 will provide the analytic framework to answer the second and third groups of questions, which will be dealt with in Sections 4 and 5 respectively. Section 6 answers the fourth group of questions and Section 7 returns to the example of Section 2 to answer the question 5. Finally, Section 8 concludes.

2. Background

In this section we try to get a clear fix on the problem under consideration. To do so, consider the following example, which is smaller and thus hopefully gives a more direct insight into what is going on than that of CS. A government funded medical research institute has four research projects on the table. These projects will reduce mortality and contribute to scientific understanding, and the company cares both about the immediate medical and broader scientific benefits. Specifically, suppose that the projects have the profile of costs and benefits shown in Table 1.

Senior management's value function over each of the two criteria is assessed as shown in Table 2, with values scores for each criterion referred to as $V_{mortality}$ and $V_{science}$ respectively.

	Reduced mortality	Scientific impact	Costs (£'000s)
a	6,000 lives saved	Excellent	200
b	4,000 lives saved	Good	100
c	4,000 lives saved	Good	100
d	3,000 lives saved	Neutral	100

Table 1: Values and costs data

	$V_{mortality}$	$V_{science}$	Costs (£'000s)
a	1	1	200
b	1/3	1/2	100
c	1/3	1/2	100
d	0	0	100

Table 2: Values and costs data

Suppose we are seeking to maximize total benefits with a budget of £200,000 and we compute the values of portfolios as the values of the constituent projects. Irrespective of the weights applied, from the data of Table 2, the most attractive portfolio seems to be $\{a\}$ which yields one unit on the scale of $V_{mortality}$, and one unit of value on the scale of $V_{science}$, whereas $\{b, c\}$ yields only 2/3 on the scale of $V_{mortality}$. At this point we might start to have doubts about the soundness of this procedure, as $\{b, c\}$ will save 8,000 people whereas $\{a\}$ will save only 6,000, and so if saving lives is weighted heavily, $\{b, c\}$ should appear more attractive.

Such doubts are reinforced if we add a project e which saves 1,000 lives, has neutral scientific impact and costs £100,000 to the project list. In this case, we rescale the values to obtain the data shown in Table 3.

	$V_{mortality}$	$V_{science}$	Costs (£'000s)
a	1	1	200
b	3/5	1/2	100
c	3/5	1/2	100
d	2/5	0	100
e	0	0	100

Table 3: Revised Values and costs

Now $\{b, c\}$ is the most highly valued portfolio whatever weights are applied. So the ordering of portfolios induced by this procedure may change when items are added to the choice set. In other words, we can observe a **rank reversal**

induced by the rescaling of the value function - a sure sign that we have done something problematic. Note that this phenomenon does not arise because of the linear transformation itself (as a monotonically increasing transformation this cannot change the ordering of projects within a criterion) but rather arises because of an interaction between the linear transformation and the use of an additive combination rule to compute overall portfolio scores. In any case, the rank reversal is a violation of the principle of Independence of Irrelevant Alternatives, according to which the preference for $\{a\}$ over $\{b, c\}$ or vice versa should not depend on whether portfolios including e are members of the choice set. Violation of this same principle has led many decision analysts to reject the Analytic Hierarchy Process as a method for supporting multicriteria choice (see e.g. Belton and Gear, 1983; Dyer, 1990; Salo and Hämäläinen, 1997).

3. Derived measurement, consistency and invariance

In the technical part of this paper, we will address the following problem: suppose you have a set of benefit numbers, and the numbers are measured on a particular scale, and the numbers are combined in a particular way to give an overall portfolio value, what does the class of admissible transformations of the underlying scale have to be in order to avoid the sort of rank reversal problem which the example demonstrates? This problem embraces the question of what goes wrong with the decision analysis procedure discussed above, but casts it in a more general setting. Practising analysts – whether decision analysts or not – have numbers (such as benefit scores), and they do operations on these numbers (such as adding them up to come up with overall portfolio values), but are they doing this in an appropriate way?

The framework we will use for studying this question is that of *derived measurement* (Suppes and Zinnes, 1963). The canonical example of derived measurement is density. The story is that someone has measured mass M and volume V by bringing some measurement technology to bear on some objects. We obtain these measurements and calculate the **derived** measurement (density) as $\frac{M}{V}$ using these original (or **fundamental**) measurements. If object 1 has mass M_1 and volume V_1 and object 2 has mass M_2 and volume V_2 we say "object 1 is denser than object 2"

6

if $\frac{M_1}{V_1} > \frac{M_2}{V_2}$. But can we be sure that if mass or volume had been measured using some different scale we would still be led to the same conclusion about relative densities?

In our approach we start with a (finite) set of objects A which are the objects of choice and a binary preference order $\succsim$ over A. When we say $\succsim$ is a preference relation we mean it is complete and transitive and so is representable by a real-valued function. The indifference relation $\sim$ and strict preference relation $\succ$ are defined in the usual way. We will be interested in the situation where we can compute a derived measurement which represents our preferences in the following sense.

Definition 1. A relation $\succsim$ on A is **represented by the function** $v : A \mapsto \mathbb{R}$ iff $a' \succsim a'' \Leftrightarrow v(a') \geq v(a'') \ \forall a', a'' \in A$.

The conditions for the existence of representing functions of various forms are well-known (Fishburn, 1970, Krantz et al, 1971).

We will suppose that the contents of A are vectors of decisions and measurements, formally that $A = \mathbb{R}^{kn} \times \{0,1\}^n$ for some positive integer n and $k \in \{1,2\}$. The intended interpretation of the $\{0,1\}^n$ is that there are n things which we could or could not do. The intended interpretation of the $\mathbb{R}^{kn}$ is that associated with each decision x_i there is a vector of measurements z_i of dimensionality equal to k. Hence a generic element of A will be written as $(z,x) = ((z_1, ..., z_i, ..., z_n), (x_1, ..., x_i, ..., x_n))$.

We restrict our focus in this paper to situations in which we are faced with binary decisions. In fact, often decisions can be thought of as polychotomous in the sense that one can choose different levels of different levels of investment (e.g. "bronze", "silver" and "gold") in a project or area. Incorporating polychotomous investment decisions would not be difficult technically but this would render the notation more cumbersome, without providing additional insights.

The specific nature of these measurements (and of the associated technology) is purposely left undefined: measurements could be directly assessed subjective values, elicited using standard decision analytic techniques (von Winterfeldt and Edwards, 1986); judgmental assessments of benefits which could be at least in principle measurable once the project is implemented using techniques drawn from

professional disciplines such as accounting or epidemiology; or actual physical measurements (e.g. in the situation where the DM is interested in purchasing scrap metal and she may be able to weigh potential purchases). We highlight that we do *not* assume that these measurements are values assigned by a value function, although they could be.

The measurement system is such that a family of permissible transformations of the numbers are possible (for example revenues may be measured in British £ and we wish to convert them into €). We call this family $\mathcal{F}$: the contents of $\mathcal{F}$ are functions $f : \mathbb{R}^{kn} \mapsto \mathbb{R}^{kn}$. By a harmless abuse of notation we will also use f to denote a function which maps $\mathbb{R}^{kn} \times \{0,1\}^n$ into itself and is defined by $f(z,x) = (f(z), x)$, and $\mathcal{F}$ to be the set of such functions. Associated with each $f \in \mathcal{F}$ (extended in this way) is a preference ordering $\mathcal{R}(f)$ on A. The indifference relation $\mathcal{I}(f)$ will be defined in the usual way. A set of functions and associated preference relations will be written $\langle \mathcal{F}, \mathcal{R} \rangle$.

Next, we define a notion of consistency.

Definition 2. A set of functions and associated relations $\langle \mathcal{F}, \mathcal{R} \rangle$ is **consistent with** $\succsim$ iff $a' \succsim a'' \Leftrightarrow f(a') \ \mathcal{R}(f) f(a'') \ \forall a' , a'' \in A, \forall f \in \mathcal{F}$.

In the ensuing "with $\succsim$" will be understood and we will just write "consistent".

Clearly this concept of consistency can be interpreted as ruling out rank reversals when we perform some operation on the project benefit scores (such as rescaling them as we did in the example of the previous section).

We wish to distinguish consistency from another condition.

Definition 3. A relation $\succsim$ is $\mathcal{F}-$**invariant** if $a' \succsim a'' \Leftrightarrow f(a') \succsim f(a'') \forall a' , a'' \in A, \forall f \in \mathcal{F}$.

Intuitively, invariance would apply (for example) if, presented with a vector of numbers representing revenues from a set of projects we could undertake, and, when asked whether it would make a material difference to our preferences over portfolios of projects whether the currency in which these revenues were denominated was \$, £, €, or ¥, we reply "no, it does not matter". This would be the case if our value function was strictly monotonically increasing, but not if we had

some particular threshold such we are indifferent to all levels of money above it. (We might find it hard to imagine what to do with the second billion € or £, but not with the second billion ¥.) Indeed, this is the difference between invariance and consistency: with consistency we are allowed to change the threshold when we rescale, with invariance not.

It is pretty clear that invariance is stronger than consistency, in the following sense:

Proposition 1. *If $\succsim$ is $\mathcal{F}-$invariant, and $\mathcal{R}(f)$ is defined by $\succsim \forall f \in \mathcal{F}$ then $\langle \mathcal{F}, \mathcal{R} \rangle$ is consistent.*

To give an example where we might have preferences which are consistent but not $\mathcal{F}-$invariant, suppose that $\mathcal{F} = \{f(z) = z - 2\}$ and $\succsim$ is defined by $z' \succsim z''$ iff $-(z'-2)^2 \geq -(z''-2)^2$. Then $2 \succ 3$ but $f(3) = 1 \succ 0 = f(2)$, so $\succsim$ is not $\mathcal{F}-$invariant. But if we define $f(z')\mathcal{R}(f)f(z'')$ by $-f(z')^2 \geq -f(z'')^2$, then $\langle \mathcal{F}, \mathcal{R} \rangle$ is consistent.

Obviously consistency is desirable from a normative point of view. Once we have converted £ into € or vice versa we should not find ourselves changing our minds about the attractiveness of our investments. However, it is also pretty weak by itself. In the ensuing we will have to supplement consistency by assuming the functions which represent $\succsim$ and $\mathcal{R}(f)$ have some particular functional form. This will make it possible to derive interesting results. Invariance, on the other hand, is a strong and normatively significant assumption, which can be used as the basis for representation theorems (see e.g. Sen, 1977).

4. One-part portfolio models

The setting for our first model is that we are confronted by $n \geq 2$ "projects" in which we can invest, and these projects have benefits if implemented. We have a measurement technology which maps the set of projects into $\mathbb{R}^n$ (i.e. $k = 1$) and an element in this space will be called $b = (b_1, ..., b_n)$. The intended interpretation of such a point is that it corresponds to a set of projects where b_i is delivered through project i. In terms of our derived measurement framework, this vector b is a fundamental measurement. If we then compute, as a derived measurement, the value of the portfolio, and we want this measurement to be consistent, what

9

properties do we require of the underlying measurement technology from which we obtain b?

The precise answer turns out to depend on the formula which is used to compute the overall portfolio value. The most natural formula is as follows.

Definition 4. If $\succsim$ on $\{0,1\}^n \times \mathbb{R}^n$ is represented by $\sum_i b_i x_i$ say it is **represented by a one-part additive portfolio model**.

The function $\sum_i b_i x_i$ is the function recommended by Kirkwood (1996), and which CS criticise as neglecting to explicitly draw attention to the value of not doing a project. The intended interpretation is that if a project i is done ($x_i = 1$) then a benefit b_i is realized.

An important class of admissible transformations is as follows:

Definition 5. The set $\mathcal{S}$ of **similarity transformations with common scaling factor** is the set $\{f(\cdot) | \exists \alpha \in \mathbb{R}_{++}, : f(b) = \alpha b \ \forall b \in \mathbb{R}^n\}$.

Obviously any similarity transformation of the underlying data will leave the preference ordering represented by $\sum_i b_i x_i$ unchanged, but are there other transformations which have this property? The answer is no.

Proposition 2. *If $\langle \mathcal{F}, \mathcal{R} \rangle$ is consistent and $\succsim$ and $\mathcal{R}(f)$ are represented by a one-part additive portfolio model $\forall f \in \mathcal{F}$, then $\mathcal{F} \subseteq \mathcal{S}$.*

PROOF. We develop the proof for the case $n = 2$: other cases are similar. We will write the representing portfolio model as $v(\cdot)$. Take an arbitrary $f \in \mathcal{F}$. First observe that because both $((b, 0), (1, 0))$ and $((b, 0), (1, 1))$ have value b:

$$((b, 0), (1, 0)) \sim ((b, 0), (1, 1)) \tag{1}$$

Now, by consistency:

$$((f_1(b), f_2(0)), (1, 0)) \mathcal{I}(f)((f_1(b), f_2(0)), (1, 1)) \tag{2}$$

Hence by representability $v((f_1(b), f_2(0)), (1, 0)) = v((f_1(b), f_2(0)), (1, 1)) \implies f_1(b) = f_1(b) + f_2(0) \implies f_2(0) = 0$. (Note that this conclusion is unaffected even if we transform $v(\cdot)$ by some strictly increasing transformation ψ; we can always

solve the resulting equation by transforming both sides by ψ^{-1} to retrieve $f_1(b) = f_1(b) + f_2(0)$). By a parallel argument $f_1(0) = 0$.

Now note that because $\succsim$ is represented by a one-part additive value function,

$$((b, 0), (1, 1)) \sim ((0, b), (1, 1)) \tag{3}$$

Hence, by consistency:

$$((f_1(b), f_2(0)), (1, 1)) \mathcal{I}(f)((f_1(0), f_2(b)), (1, 1)) \tag{4}$$

Hence, by representability $f_1(b) + f_2(0) = f_1(0) + f_2(b) \implies f_1(b) = f_2(b) \forall b$. Henceforth write ϕ for both f_1 and f_2.

Now observe that:

$$((b_1, b_2), (1, 1)) \sim ((b_1 + b_2, 0), (1, 1)) \tag{5}$$

Hence, by consistency again:

$$((\phi(b_1), \phi(b_2)), (1, 1)) \mathcal{I}(f)((\phi(b_1 + b_2), \phi(0)), (1, 1)) \tag{6}$$

From which we can conclude that $\phi(b_1) + \phi(b_2) = \phi(b_1 + b_2)$, which is a Cauchy functional equation. As is well-known (Aczél, 2006), the only solution to this equation is $\phi(b) = \alpha b$. Hence the result is proved.

Recall that an absolute scale is one where no transformations are permissible; a ratio scale is one where similarity transformations (ie multiplication by a scaling factor) are permissible; and an interval scale is one where affine transformations (multiplication by a scaling factor and addition of a positive constant) are permissible. So the implication of Proposition 2 is that if the project benefit scores are measured on an absolute scale, no rank reversal is possible; if they are measured on a ratio scale, no rank reversal is possible; but if they are measured on (for example) an interval scale, then a rank reversal can occur, indicating that we have a problem. Value functions are generally taken to be measured on an interval scale, and it is this which justifies the rescaling as the option set changes, which in turn gives rise to the rank reversal described in Section 2.

It is important to realise that the above analysis applies only to settings where we are making *portfolio* choices. In settings where we are choosing one item from a set, defining value functions in this way and choosing the item with the highest value is unproblematic advice. It is also important to realise that the set of

consistent transformations (ie transformations which do not induce a rank reversal) is dependent on the choice of portfolio model. For example, if the portfolio model is $\max_i\{b_i x_i\}$ then $\mathcal{F}$ can contain any strictly increasing function without fear of inducing an inconsistency (the best portfolios are those which contain the best project, and no strictly increasing transformation will change the ranking of the projects), and so rescalings are unproblematic, as long as the benefits of all projects are rescaled by the same amount. To take another example, if the portfolio model is $\Pi_i b_i x_i$ then $\mathcal{F}$ can be contained in the following set without introducing an inconsistency:

Definition 6. The set $\mathcal{W}$ of weighted power transformations is the set $\{f(\cdot)|\exists \alpha_i \in \mathbb{R}_{++}, \exists p \in \mathbb{R}_{++} : f_i(r) = \alpha_i r^p \ \forall r \in \mathbb{R}\}$.

In other words, the range of transformations which may feature in $\mathcal{F}$ without inducing an inconsistency is enlarged - but nevertheless, measuring benefits on interval scales which permit a change in zero point (for example, value functions) is still problematic even with this choice of one-part additive portfolio model.

5. Two-part portfolio models

In the second formalisation we have $n \geq 2$ "sectors" in which we can invest or not. Investing in a sector leads to different outcomes than if we had not invested. Sectors could be locations, or diseases, or physical installations, such as roads or flood defences; if one is planning flood defences, the decision could be to upgrade from sand defences to concrete. We have a measurement technology which measures the outcomes associated with investing and not investing, i.e now our measurement technology produces $2n$ measurements ($k = 2$). Specifically, associated with each sector i we have a measurement of the outcome $\overline{z}_i$ in the case where we invest and $\underline{z}_i$ in the case where we do not invest. Write $z_i = (\overline{z}_i, \underline{z}_i)$. These are our fundamental measurements in this setup. Again we are agnostic about what these outcomes are: they could be subjectively assessed values or wealth levels, number of premature deaths from illness, or expected monetized damage from flooding.

We use a model of the form $\sum_i \big(\overline{z}_i x_i + \underline{z}_i(1 - x_i)\big)$ as a representing portfolio function for these preferences (this is the model which is recommended in CS as equation (3)). It will be noted that when this model $\sum_i \big(\overline{z}_i x_i + \underline{z}_i(1 - x_i)\big)$ is used the conclusions of the previous section do not hold: in particular, a transformation of the form $f_i(z_i) = z_i + \beta$ applied to both $\overline{z}_i$ and $\underline{z}_i$ will obviously leave the ranking of solutions induced by this portfolio model unchanged.

Definition 7. If $\succsim$ is represented by $\sum_i \big(\overline{z}_i x_i + \underline{z}_i(1 - x_i)\big)$ say it is **represented by a two-part additive portfolio model**.

We define the following set of transformations. Note that α is a scalar and β is a vector.

Definition 8. The set $\mathcal{A}$ of **affine transformations with a common scaling factor** is the set $\{f(\cdot)|\exists \alpha \in \mathbb{R}_{++}, \exists \beta \in \mathbb{R}^n : f(z) = \alpha z + \beta \forall z \in \mathbb{R}^n\}$.

As in the previous section, the transformations in the designated set are the only transformations to leave the preference ordering unchanged.

Proposition 3. *If $\langle \mathcal{F}, \mathcal{R} \rangle$ is consistent and $\succsim$ and $\mathcal{R}(f)$ are represented by a two-part additive portfolio model $\forall f \in \mathcal{F}$, then $\mathcal{F} \subseteq \mathcal{A}$.*

PROOF. Take an arbitrary $f \in \mathcal{F}$. Observe that the indifference $((z,0),(0,0),(1,0)) \sim ((0,0),(z,0)(0,1))$ follows from representability. Hence from consistency:

$$(((f_1(z), f_1(0)),(f_2(0), f_2(0))),(1,0))\,\mathcal{I}(f)\,(((f_1(0), f_1(0)),(f_2(z), f_2(0))),(0,1))) \tag{7}$$

from this it follows that:

$$f_1(z) + f_2(0) = f_1(0) + f_2(z) \tag{8}$$

or in other words, $f_1(z)$ and $f_2(z)$ differ only in a constant term. Call this term κ and write $f_2(z) = f_1(z) + \kappa$.

Now observe that $(((z_1, 0),(z_2, 0)),(1,1)) \sim (((z_1 + z_2, 0),(0,0)),(1,1))$ and hence:

$$(((f_1(z_1), f_1(0)),(f_1(z_2) + \kappa, f_1(0) + \kappa)),(1,1))\,\mathcal{I}(f)$$
$$(((f_1(z_1 + z_2), f_1(0)),(f_1(0) + \kappa, f_1(0) + \kappa)),(1,1)) \tag{9}$$

13

Because the κs cancel, it follows that:

$$f_1(z_1) + f_1(z_2) = f_1(z_1 + z_2) + f_1(0) \tag{10}$$

This is equivalent to:

$$f_1(z_1) - f_1(0) + f_1(z_2) - f_1(0) = f_1(z_1 + z_2) - f_1(0) \tag{11}$$

Hence we can define $\gamma(z) = f_1(z) - f_1(0)$ and obtain the following functional equation:

$$\gamma(z_1) + \gamma(z_2) = \gamma(z_1 + z_2) \tag{12}$$

This is the Cauchy equation again, so $\gamma(z) = \alpha z$. Write $\beta_1 = f_1(0)$ and $\beta_2 = f_1(0) + \kappa$ and we have an affine transformation with common scaling factor as claimed.

Hence, for these two part functions, using value functions (or other measurement technologies) which provide interval valued measurements, does not represent a threat to consistency and place us at risk of inducing a rank reversal.

Does this mean that we can disregard nature of the scales on which the outcomes are measured when using such two-part models? Absolutely not! Consider the following example.

Example 1. *Suppose outcomes are measured on an ordinal scale, and portfolios are evaluated by a two-part additive model. Consider the following data shown in Table 4 for some evaluation problem.*

	$\overline{z}$	$\underline{z}$	Costs (£'000s)
a	3	1	100
b	3	1	100
c	4	1	200

Table 4: Outcomes and cost data

Suppose we have a budget of £200,000. The optimal choice is portfolio $\{a, b\}$ because the value of the model $\sum_i \overline{z}_i x_i + \sum_i \underline{z}_i (1 - x_i)$ is $3 + 3 + 1 = 7$ whereas if we choose $\{c\}$ we obtain $1 + 1 + 4 = 6$. However, if a transformation $f(z) = z^3$ is applied to the $\overline{z}$s and $\underline{z}$s (since all strictly increasing transformations of an ordinal scale are permissible), then the data become as shown in Table 5.

Now if we apply the formula $\sum_i f(\overline{z}_i) x_i + \sum_i f(\underline{z}_i)(1 - x_i)$ $\{a, b\}$ is associated with an value of $27 + 27 + 1 = 55$ and $\{c\}$ with a value of $64 + 1 + 1 = 66$ so the order has reversed.

14

	$f(\overline{z})$	$f(\underline{z})$	Costs (£'000s)
a	27	1	100
b	27	1	100
c	64	1	200

Table 5: Transformed outcome and costs data

Comparing Proposition 2 and Proposition 3 gives us some insight into why using value functions (or indeed any measurement technology which produces numbers measured on an interval scale) to evaluate benefits is problematic in the case of a one-part additive portfolio model but non-problematic in the case of a two-part additive portfolio model. However, Example 1 shows us that we would experience a problem exactly analogous to the problem of Section 2 when using a two part function additive portfolio model if we were to use a technology which produced measurements on an ordinal scale to measure benefits.

Moreover, our positive conclusion that it is safe to choose an interval scale when using a two-part portfolio model is not necessarily good if the model in question does not have an additive form - for example a model with the multiplicative form $\prod_i \big(\overline{z}_i x_i + \underline{z}_i(1 - x_i)\big)$. Consider the following example.

Example 2. *We are seeking to build a small research team, and we need a mathematician (m) and a programmer (p). We have people with both these skills in house, but we have some funds and can hire either a better m or a better p (but not both). We rate both existing and potential new hire ms and ps on a zero to 100 scale as follows, with $\overline{z}$ denoting the score for the new hire and $\underline{z}$ denoting the score for the existing staff, as in Table 6.*

	$\overline{z}$	$\underline{z}$
m	48	30
p	60	40

Table 6: Original staff evaluations

We believe that mathematics and programming are complementary skills and so we combine scores multiplicatively to arrive an overall score for the team. Within our budget constraint, we have two possibilities – the portfolio containing the new m and the existing p with a total score of $48 \times 40 = 1920$ or the portfolio containing the existing m and new p with a total score of $30 \times 60 = 1800$. So clearly choosing to hire the new mathematician is the way to go. However, we have a second

15

thought - these are skilled professionals, so surely we should start the scale at 100 instead of zero and go up to 200? We thus transform our scores by the function $g(z) = z + 100$. This gives us new scores as shown in Table 7.

	$g(\overline{z})$	$g(\underline{z})$
m	148	130
p	160	140

Table 7: Transformed staff evaluations

Now we recompute the portfolio scores. This time, the portfolio containing the new m and the existing p has a total score of $148 \times 140 = 20720$ and the portfolio containing the existing m and new p has a total score of $130 \times 160 = 20800$, so it seems now we should hire the new programmer instead.

In this case we have performed an affine transformation of the underlying scale – quite permissible if the scale is indeed an interval scale – and have obtained a rank reversal, even though we are using a two-part model. This is not a coincidence, and in fact it can be shown (using the same line of reasoning as that of Propositions 2 and 3) that in the case of a two-part multiplicative model the only transformations of the underlying scales which preserve consistency are those in the set $\mathcal{W}$, as introduced in the last section.

6. Beyond separable models

So far in this paper, we have identified admissible transformations of underlying data which preserve particular value models. However, a possible objection to this analysis is that we have so far focussed exclusively on additive and other separable value models. This is a significant restriction in a portfolio context as often there will be interactions between decisions about particular projects or particular sectors. For example, the value of investing in secondary care facilities for people suffering from stroke or advanced kidney disease may depend on whether one has also invested in preventive activity for these conditions; to the extent that one has not invested in prevention, one can expect secondary care facilities to have higher utilization and thus add more value.

An interesting instance of this sort of model is described in Mild and Salo (2009). These authors describe an application undertaken for the Finnish Road

Administration where particular profiles of maintenance decisions are mapped into a descriptive system which captures the distribution of the various elements of the road infrastructure across a number of quality classes. This in turn is mapped by a multiattribute value function with four criteria representing road safety, asset value preservation, customer satisfaction and environmental aspects, into a value score. In our terms, one can think of this formal apparatus as a device for mapping a profile of decisions into an outcome space, with non-zero outcomes associated both with choosing both with "doing" and "not doing" a particular piece of maintenance. Such a model "feels like" a two-part model but the outcome values need not combine additively over sectors (for example, in the Mild and Salo (2009) model, as there is a Markov chain which governs the degradation of the roads over time, a maintenance action in one period may interact with a maintenance action in a subsequent period).

Is there anything we can say, then, about these situations where we do not have separability between decisions? It turns out there is. To see what may be generalizable to this more complex environment, make the following definition:

Definition 9. The set $\mathcal{T}$ of **translations** is the set $\{f(\cdot) | \exists \beta \in \mathbb{R}^n, : f(b) = b + \beta \ \forall b \in \mathbb{R}^n\}$.

One of the striking feature of the previous section has been that the project value models of Section 4 are not consistent with respect to sets of data transformations containing $\mathcal{T}$ (even the portfolio model $\max_i \{b_i x_i\}$ requires us to transform each project score by the same monotonic function), whereas the sector value models of Section 5 may be (for example the class of consistent transformations of the two part additive model is a superset of $\mathcal{T}$).

To see how this insight applies in a more general setting, in this section, rather than studying relations represented by particular functions, we study a property imposed on the relation itself directly. We continue to suppose that $\succsim$ is a preference ordering but no longer suppose that it has an explicit representation through a specific portfolio model. Instead we will be interested in the following property.

Definition 10 (PC). A preference relation $\succsim$ over $A = \mathbb{R}^{kn} \times \{0,1\}^n$ ($k \in \{1,2\}$) will be said to be **one-part coherent** if (i) $(\mathbf{0}^{kn}, \mathbf{1}^n) \sim (\mathbf{0}^{kn}, \mathbf{0}^n)$ and (ii) $(z', \mathbf{0}^n) \sim$

$(z'', \mathbf{0}^n) \forall z', z'' \in \mathbb{R}^{kn}$ where $\mathbf{0}^{kn}, \mathbf{0}^n$ are the zero vectors of dimensionality kn and n respectively, and $\mathbf{1}^n$ is a vector of ones of dimensionality n.

Thus, part (i) of one-part coherence captures the idea that a zero on the benefit scale has the meaning that one is indifferent between doing a project and not doing it. Part (ii) says that if one does not do any project, it does not matter how much benefit would have been realized had one done some.

It is interesting to compare with this what one may think of as the essential features of the two-part representation. Consider the following axiom.

Definition 11 (SC). A preference relation $\succsim$ over $A = \mathbb{R}^{2n} \times \{0,1\}^n$ will be called **two-part coherent** if

$$((\bar{z}_1, \underline{z}_1), ..., (z_i, \underline{z}_i), ..., (\bar{z}_n, \underline{z}_n)), (x_1, ..., 1, ..., x_n))$$
$$\sim ((\bar{z}_1, \underline{z}_1), ..., (\bar{z}_i, z_i), ..., (\bar{z}_n, \underline{z}_n)), (x_1, ..., 0, ..., x_n))$$

$\forall i \in \{1, ..., n\}$, $\forall (\bar{z}_1, \underline{z}_1), ..., (\bar{z}_{i-1}, \underline{z}_{i-1}), (\bar{z}_{i+1}, \underline{z}_{i+1}), ..., (\bar{z}_n, \underline{z}_n)) \in \mathbb{R}^{2(n-1)}$ and $\forall \bar{z}_i, z_i, \underline{z}_i \in \mathbb{R}$.

This axiom captures the idea that whether one obtains z_i from choosing 1 or from choosing 0 is a matter of indifference: it is the same z_i in either case.

We also need a very mild axiom.

Definition 12. A preference relation $\succsim$ over $A = \mathbb{R}^{kn} \times \{0,1\}^n$ will be said to be **essential** if $\exists z', z'' \in \mathbb{R}^{kn} : (z', \mathbf{1}^n) \succ (z'', \mathbf{1}^n)$.

To see that one-part and two-part coherence are very different ideas, consider the following result.

Proposition 4. *No preference relation $\succsim$ over $A = \mathbb{R}^{2n} \times \{0,1\}^n$ can be both one-part coherent, two-part coherent and essential.*

PROOF. For conciseness, write $z = [(\bar{z}_i, \underline{z}_i)]$. Choose $z' = [(\bar{z}'_i, \underline{z}'_i)], z'' = [(\bar{z}''_i, \underline{z}''_i)] :$ $(z', \mathbf{1}^n) \succ (z'', \mathbf{1}^n)$. By part (ii) of **PC**, $([(z'_i, \bar{z}'_i)], \mathbf{0}^n) \sim ([(z''_i, \bar{z}''_i)], \mathbf{0}^n) \, \forall z'_i, z''_i \in \mathbb{R}$. By **SC**, $([(z'_i, \bar{z}'_i)], \mathbf{0}^n) \sim([(\bar{z}'_i, \underline{z}'_i)], \mathbf{1}^n)$ and $([(z''_i, \bar{z}''_i)], \mathbf{0}^n) \sim([(\bar{z}''_i, \underline{z}''_i)], \mathbf{1}^n)$. From the transitivity of indifference, we can conclude $(z', \mathbf{1}^n) = ([(\bar{z}'_i, \underline{z}'_i)], \mathbf{1}^n) \sim([(\bar{z}''_i, \underline{z}''_i)], \mathbf{1}^n) = (z'', \mathbf{1}^n)$ which is a contradiction to essentiality.

We are able to prove an impossibility theorem for one-part coherence.

Theorem 3. *For all preference relations $\succsim$ over $A = \mathbb{R}^n \times \{0,1\}^n$, if $\langle \mathcal{F}, \mathcal{R} \rangle$ is consistent and $\succsim$ and $\mathcal{R}(f)$ are essential and one-part coherent $\forall f \in \mathcal{F}$ then $\mathcal{T} \nsubseteq \mathcal{F}$.*

PROOF. Suppose $\mathcal{T} \subseteq \mathcal{F}$ and consider z' and z'' distinct such that $(z', \mathbf{1}^n) \succ (z'', \mathbf{1}^n)$. Now apply the transformation $f_i(z_i) = z_i - z_i'$ $\forall i$ to z' to obtain: $((f_1(z_1), ..., f_n(z_n)), \mathbf{1}^n) = (\mathbf{0}^n, \mathbf{1}^n)$. By **PC** part (i), $(\mathbf{0}^n, \mathbf{1}^n) \mathcal{I}(f)(\mathbf{0}^n, \mathbf{0}^n)$ and so by consistency, $(z', \mathbf{1}^n) \sim (z', \mathbf{0}^n)$. By parallel reasoning, $(z'', \mathbf{1}^n) \sim (z'', \mathbf{0}^n)$. Hence by transitivity of strict preference, $(z', \mathbf{0}^n) \succ (z'', \mathbf{0}^n))$. But this is a contradiction to **PC** part (ii).

Thus, it is not possible to measure the benefit values in any one-part model on a scale which permits arbitrary translation without exposing oneself to the possibility of rank reversal. Contrast this with the main result (Proposition 3) which shows that it is possible to measure the benefit values in two-part models on such a scale with no such risk.

The non-permissibility of translation in portfolio models which represent project-wise coherent preferences is important for practical purposes, in that it shows that the reason for the sorts of pathologies by the one-part model exhibited in Section 2 do not arise because of some quirk of the additive form: in fact they are much more general. Moreover, as will be apparent in the next section, the idea of one-part coherence gives actual operational guidance into how a natural baseline for measurement can be fixed to avoid the sorts of rank reversal pathologies documented in Section 2.

7. Implications for practice

In this section, we return to the fundamental question: how should practising analysts model problems like that of Section 2? CS are proponents of the additive two-part model $\sum_i \overline{z}_i x_i + \sum_i \underline{z}_i (1 - x_i)$ as a way of dealing with the baseline setting issue. According to these authors "it is best to consider each project individually by explicitly scoring not doing the project on each attribute and using these scores to calculate an overall project-specific score, v_i^o [their notation], for not doing project i." (p260). Somewhat *contra* this view, we argue that there are cases

where it is more natural to use one-part models. We illustrate our argument by revisiting the case where a government funded medical research institute wishes to select a portfolio of projects without spending in excess of a budget. The aim is to develop medications for a potentially lethal disease (let us say diabetes). One of the criteria for evaluation is impact on mortality. We argue that it is more natural to use a one-part model rather than a two-part model in this case.

The principal reason is that a significant attractive feature of the one-part model $\sum_i b_i x_i$, is that one does not have to explicitly to identify the outcomes in terms of how many people die from diabetes associated with not doing a project, which may be contentious and not decision relevant. For example, knowing the precise extent of diabetes in a population (as long as it is agreed that there is "a lot") may not be relevant to decisions between a number of "small" projects to redress diabetes, since only once the projects become sufficiently large that one may run out of people to treat does the scale of the underlying health problem become an issue. This is a real concern in economically underdeveloped countries with poor systems of public health surveillance and death registration, estimating how many people in the general population have a given disease and how many people die from that disease to a high level of accuracy is for all practical purposes impossible.

Suppose one agrees with these two reasons that it is more natural to implement a *one-part model scenario* but wishes to assess value functions over the benefits (expressed in terms of reduction in diabetes mortality), with a view to trading off mortality reduction against other criteria. We know from the example of Section 2 and Proposition 2 that we should avoid using interval scales so as not to expose ourselves to the possibility of rank reversal. To do this we need to find a natural baseline for measurement of the value function which will not vary as we change as we introduce new projects. In the spirit of the notion of "one-part" coherence of the previous section, the natural baseline for measurement is "zero projects implemented" which in this case means "zero lives saved". Note the contrast between the CS suggestion and our suggestion - we are proposing that the baseline for measurement of the benefit of a project should not be the project-specific concept of "not doing a [ie the given] project" but rather the general concept of "not doing any projects from the choice set at all".

20

To illustrate how this might work, let us return to the example of Section 2. If one evaluates "lives saved" relative to this "zero projects implemented"/ "zero lives saved" baseline as the zero of the measurement scale, takes the level of the most beneficial project - "6,000 lives saved"- as the unit and assumes a linear value function, referring back to the original column of Table 1, one obtains value scores of 1, 2/3, 2/3 and 1/2 for the different projects respectively. Note that this value measurement system does not give unique measurements: if we add an extra project which saves 10,000 lives to the option set, we have to rescale the value scores to get 3/5, 2/5, 2/5 and 3/10 for a,b,c,d respectively, but this simply involves multiplying the original value scores by a common constant (a similarity transformation), so the ranking of portfolios we obtain from summing up these value scores is unaffected. Proposition 2 tells us that whenever we want to add up value scores in this way, we can only be sure that our measurement system will not induce a rank reversal of portfolios if the permissible score transformation are similarity transformations. This is the reason that the original procedure of Section 2 gets us into difficulties. Now, because we have chosen a baseline which is independent of the choice set and does not change when we add projects, we no longer have to worry about rank reversals.

As a matter of fact, this approach of setting a so-called "do nothing" baseline has been used widely in the multicriteria context in practice, in papers as early as e.g. Bana e Costa et al. (1999, 2002). Indeed, this is what is done in the case study presented in Phillips and Bana e Costa (2007) (despite these authors' advice to set the 0 at the level of the least preferred option). Hence we see the contribution of this paper as being in part to justify the procedure of these authors, which recognises the problem identified by CS but approaches its solution in a different and to our way of thinking more natural manner.

Hence, while we recognise the issue which CS raise, our view is that the solution that they suggest - using $\sum_i \overline{z}_i x_i + \sum_i \underline{z}_i (1 - x_i)$ - is not always the most natural modelling solution. It is perfectly possible to continue to describe problems using the additive one-part model $\sum_i b_i x_i$ as long as one takes care the the coefficients are measured in a way so that they are unique at least up to rescaling by a fixed constant, and a natural way to do this is to by using a do-nothing baseline

21

for measurement which is independent of the set of projects being considered for inclusion in the project portfolio. As part of the theoretic contribution of this paper, our "one-part coherence" axiom of Section 6 can give insight into the most appropriate way to identify this baseline.

8. Conclusion

This paper has studied the issue of the choice of measurement scales to measure project benefits in multicriteria portfolio decision analysis using a derived measurement frame, motivated by an issue in MCDA practice. The distinctive feature of our frame here has been that we suppose that we have some measurements, and some way of combining them, and wish the resulting derived measurement to have certain consistency properties. We then ask what properties the underlying technology which has produced our measurements would have to have (bearing in mind that these measurements may be from judgemental elicitation, from the use of the techniques of professional disciplines such as accounting or epidemiology, or from some actual physical measurement system).

We relate our conclusions back to our motivating questions as follows:

1. *Can we get a clear view of what goes wrong when we use value functions to value projects multicriteria portfolio decision analysis?* We provide a small toy example showing that, in an environment where additive functions are used to combine project values, rescaling value functions by changing the baseline of measurement can lead to rank reversals between portfolios.

2. *What are the specific characteristics of value functions which give rise to this problem? Might this problem apply more generally when project benefits are assessed in some other way? In what way is this a specifically portfolio problem?* As long as the one-part additive portfolio model is used to transform project benefit values to portfolio scores, benefits have to be measured on a scale on which the only permissible transformations are similarity transformations if rank reversals are to be avoided, but value functions are generally taken to be interval scales: this is their problematic characteristic. This is a specifically portfolio problem in the sense that whether using a particular

scale for the measurement of benefits leads to a rank reversal problem depends on the nature of the portfolio model used to combine benefit scores - for example one could safely use interval scales for the measurement of benefit functions if a max function was use to transform value scores.

3. *Why does the solution proposed by CS "work" to resolve this problem? Is the solution proposed by CS vulnerable to other similar and related problems?* In the alternative "two-part" additive model which CS propose, the class of transformations by which the benefits can be rescaled without inducing a rank reversal is the class of affine transformations, so using value functions is unproblematic in this context. But rank reversals can still be induced in the two two-part additive models if benefits are assessed on other scale types (e.g. ordinal) and rescaling project values by an affine transformation is still a problem if a two-part multiplicative model is used – so two part models are not immune to rank reversal problems if one is careless about the nature of the scale on which the underlying benefits are measured.

4. *How important is the separability of the rule for combining project benefit values as overall portfolio benefits in this analysis? Do the results we obtain for this rule give insights into what should happen in a more general setting?* Even in the case of non-separable portfolio models, there is a clear difference between the one- and two-part models, in that the one-part models can never be used with benefit scores assessed on scales which allow arbitrary scale translation, whereas this is unproblematic in the case of the two-part additive model.

5. *What, practically speaking, can we learn from this analysis about how to perform better and more robust analysis?* We argue *contra* CS that one-part models often provide a more natural frame for modelling, rather than two-part models which require one to separately define and evaluate the outcomes associated with each and every project. We show that our axiom of one-part coherence provides a theoretically based way to identify a "do-nothing baseline" for value measurement. Consistent use of this baseline rather than a shifting baseline which changes when the option set changes is a natural way to establish a value scale with the necessary properties to be used in portfolio analysis without the possibility of rank reversal.

23

Our philosophy in writing this paper has been that mistakes (like the mistake highlighted by CS) are worth studying in some detail as a close understanding of how one has gone wrong in the past can ensure that one learns the right lessons and does not overlearn (for example it is not the case that it is *always* wrong to use interval scales to assess benefit values when using one-part models), and avoid making similar mistakes in the future (for example using interval scales in the setting where one has two-part multiplicative models). We hope that the discussion and findings in this paper will help researchers and practitioners draw appropriate conclusions about how to assess benefits in portfolio settings.

References

1. Aczél, J. (2006). *Lectures on functional equations.* Mineola, NY: Dover.

2. Baltussen, R. and Niessen, L. (2006) Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost-effectiveness and resource allocation,* 4(14).

3. Bana e Costa, C., da Costa-Lobo, M. L., Ramos, I. A., & Vansnick, J.-C. (2002). Multicriteria approach for strategic planning: the case of Barcelos. In D. Bouyssou, E. Jacquet-Lagrèze, P. Perny, R. Słowinsky, D. Vanderpooten & P. Vincke (Eds.), Aiding decisions with multiple criteria: essays in honour of Bernard Roy: Kluwer.

4. Bana e Costa, C. A., Ensslin, L., Correa, E. C., & Vansnick, J.-C. (1999). Decision Support Systems in Action: Integrated Application in a MultiCriteria Decision Process. *European Journal of Operational Research*, 113(2), 315-335.

5. Belton, V., & Gear, T. (1983). On a Short-Coming of Saaty Method of Analytic Hierarchies. *Omega-International Journal of Management Science*, 11(3), 228-230.

6. Belton, V., & Stewart, T. J. (2002). *Multiple Criteria Decision Analysis: an integrated approach.* Boston, MA: Kluwer.

7. Clemen, R. T., & Smith, J. E. (2009). On the choice of baselines in multiattribute portfolio analysis: a cautionary note. *Decision Analysis*, 6(4), 156-262.

8. de Almeida, A.T., Vetschera, R. (2012). A note on scale transformations in the PROMETHEE V method. *European Journal of Operational Research*, 219, 198-200.

9. de Almeida, A.T., Vetschera, R., Almeida J. A. (2014). Scaling Issues in Additive Multicriteria Portfolio Analysis. In F. Dargam, J. E. Hernández; P. Zaraté. S. Liu, R. Ribeiro, B. Delibasic, J. Papathanasiou (Eds.),Decision Support Systems III - Impact of Decision Support Systems for Global Environments. LNBIP 184 (Lecture Notes in Business Information Processing): Springer. 131-140.

10. Dyer, J. S. (1990). Remarks on the Analytic Hierarchy Process. *Management Science*, 36(3), 249-258.

11. Fishburn, P. C. (1970). *Utility theory for decision making.* Chichester: Wiley.

12. Golabi, K., Kirkwood, C. W., & Sicherman, A. (1981). Selecting a portfolio of solar energy projects using multiattribute preference theory. *Management Science*, 27, 174-189.

13. Kirkwood, C. (1996). *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets.* Belmont, CA: Duxbury.

14. Kleinmuntz, C. E., Kleinmuntz, D. N. (1999). A strategic approach to allocating capital in health-care organizations. *Healthcare Financial Management* 53(4) 52–58.

15. Kleinmuntz, D. N. (2007). Resource allocation decisions. In W. Edwards, R. F. Miles & D. von Winterfeldt (Eds.), Advances in decision analysis (pp. 400-418). Cambridge: CUP.

16. Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). Foundations of Measurement Vol 1. New York: Academic Press.

17. Liesiö, J. , Punkka, A. (2014). Baseline value specification and sensitivity analysis in multiattribute project portfolio selection. *European Journal of Operational Research* 237, 946–956.

18. Lindstedt, M., Liesiö, J. , Salo, A. (2008). Participatory development of a strategic product portfolio in a telecommunication company. *International Journal of Technology Management,* 42(3), 250–266.

19. Mild, P., & Salo, A. (2009). Combining a Multiattribute Value Function with an Optimization Model: An Application to Dynamic Resource Allocation for Infrastructure Maintenance. *Decision Analysis,* 6(3), 139–152.

20. Milnor, J. (1954). Games against nature. In R. M. Thrall, C. M. Coombs & R. L. Davis (Eds.), Decision Processes. New York: Wiley.

21. Morcos, M. S. (2008). Modelling resource allocation of R&D project portfolios using a multi-criteria decision-making methodology. *International Journal of Quality and Reliability Management,* 25(1), 72-86.

22. Morton, A. (2010). Bridging the gap: health equality and the deficit framing of health. *Health Economics*, 19(12), 1497-1501.

23. Morton, A., & Fasolo, B. (2009). Behavioural Decision Theory for Multi-Criteria Decision Analysis: a guided tour. *Journal of the Operational Research Society*, 60(2), 268-275.

24. Parnell, G. S., Bennett, G. E. , Engelbrecht, J. A., Szafranski, R. (2002). Improving resource allocation within the National Reconnaissance Office. *Interfaces,* 32(3), 77–90.

25. Phillips, L. D., & Bana e Costa, C. (2007). Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research*, 154(1), 51-68.

26. Rao, V. R., & Weiss, E. N. (1986). On Resource-Allocation Problems with Interval-Scale Coefficients. *Journal of the Operational Research Society*, 37(6), 631-635.

27. Roberts, F. S. (1990). Meaningfulness of conclusions from combinatorial optimization. *Discrete Applied Mathematics*, 29(2-3), 221-241.

28. Salo, A. A., & Hämäläinen, R. P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, 6(6), 309-319.

29. Salo, A. A. , Keisler, J., Morton, A., eds. (2011). *Portfolio Decision Analysis: Methods for improved resource allocation.* New York: Springer.

30. Sen, A. (1977). On Weights and Measures - Informational Constraints in Social-Welfare Analysis. *Econometrica*, 45(7), 1539-1572.

31. Suppes, P., & Zinnes, J. L. (1963). Basic Measurement Theory. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), Handbook of Mathematical Psychology. New York: Wiley.

32. von Winterfeldt, D., W. Edwards. (1986). *Decision analysis and behavioral research.* Cambridge: CUP.