Constructing Meaningful Sustainability Indices*

Heinz Welsch

University of Oldenburg, Department of Economics, 26111 Oldenburg, Germany, welsch@uni-oldenburg.de

Keywords. *Environmental Index, Sustainability Index, Preference Ordering, Comparability, Measurability, Environmental Indicator*

JEL. *C4; Q00*

Abstract. This paper surveys and evaluates the possibilities and limitations of sustainability indices from the point of view of meaningfulness. A sustainability index is defined as meaningful if it allows unambiguous orderings of the relevant 'situations' over time independent of the measurement units in which the variables describing the situations are expressed. The cases of commensurability and incommensurability are distinguished. In the former, the comparison of situations is unambiguous be-cause all legitimate choices of measurement units can be accommodated on the basis of exogenously given relationships among the variables. These relationships may define a monetary welfare-metric or a bio-physical effects-metric. In the case of incommensurability, common approaches (both cardinal and ordinal) may fail to yield meaningful indices. A systematic assessment of which indices are meaningful in which circumstances is provided.

1 Introduction

"Don't run down your assets!" – The sustainability imperative can be put as simple as that. There are, however, a variety of assets that may be worth preserving: natural capital, man-made (physical) capital, human capital, not to speak of 'social capital' (governance, trust, and other social institutions). Different notions of sustainability differ with respect to the degree of substitutability which is presumed to exist between the various types of capital. A hypothetical extreme position might entail that each and every asset should be preserved: Not only should the stocks of natural capital, physical capital, human and social capital be non-decreasing but also the different kinds of natural capital, down to individual species, minerals, or fuels.

* Acknowledgement: I am grateful to Udo Ebert, Friedrich Schneider, Armin Schmutzler, Christoph Weber, Cees Withagen and Pascal Wolff for useful comments on an earlier version.

Such a position could be called 'ultra-strong sustainability'. It has a big advantage: To be monitored, it does not need the construction of any 'sustainability index' whatsoever. But ultra-strong sustainability is not a tenable position in the real world, be it only since it would imply that all non-renewable resources remain untouched indefinitely. By contrast, both analysts and policy makers will normally be prepared to tolerate some trade-off between different assets, and this begs the need for sustainability indices, that is, tools for answering the question: "Have the relevant assets been kept intact *overall*?" (weak sustainability of some degree).

Speaking somewhat loosely, a sustainability index should permit an assessment of whether 'the situation' (e.g. the environmental situation) has become better or worse between time t and $t+1$. This *sustainability problem* is slightly different from the *ranking problem*: How do places (e.g. countries) rank in terms of 'the situation' in question? Though both problems are related, I will mainly focus on the former, touching upon the latter only occasionally.

A basic requirement when constructing a sustainability index is that it should be meaningful, in the sense that the comparison of situations over time should be unambiguous with respect to the choice of measurement units of the relevant variables.

With respect to meaningfulness it is useful to distinguish between the case of commensurability and the case of incommensurability. In the former case, the comparison of situations is unambiguous because all legitimate choices of measurement units can be accommodated on the basis of exogenously given relationships among the variables. This is not the case with incommensurable variables: Here an ambiguity problem may arise, depending on the measurability and comparability properties of the variables involved. The two cases will be addressed in separate sections (Section 2, and Sections 3 and 4, respectively).

The focus of the paper is on methodological issues. Indices actually proposed or applied are mentioned mainly for illustrative purposes. A comprehensive survey of actual indices is not intended.

2 Commensurability

Two types of sustainability indices considered in the literature fall into the category of commensurability: indices based on a monetary welfare metric and indices based on a bio-physical effects-metric, respectively.

2.1 Monetary Welfare-Metric

To illustrate the welfare-based approach to constructing sustainability indices, consider a welfare function defined over consumption C and the stock of natural capital N. Natural capital is an aggregate which comprises various types of exhaustible and renewable resources as well as various dimensions of environmental quality. The welfare function reads

$$W_t = W(C_t, N_t) \tag{1}$$

and is assumed to be twice differentiable, increasing and strictly concave in both arguments. The variable W_t is a monetary measure of per-capita welfare or utility at time t.[1]

A definition of sustainability might involve the following requirements:

(a) $\dot{W}_t \geq 0$ for all t, (b) $W_t > 0$ for all t, that is, welfare should be non-declining and positive. If we accept this definition, W_t is a sustainability index. It can be linked to other, perhaps more popular, notions and indices of sustainability as follows. Consider a well-behaved (per-capita) production function $Y(K,R)$, where K denotes the stock of physical capital and R the flow of natural resources into the production process. Letting S denote savings, we have $C = Y - S$, and requirement (a) can be written as follows (with subscripts denoting partial derivatives and dots denoting time derivatives):

$$\begin{aligned} \dot{W} &= W_C \cdot (Y_K \dot{K} + Y_R \dot{R} - \dot{S}) + W_N \dot{N} \\ &= W_C Y_K \dot{K} + W_N \dot{N} + W_C \cdot (Y_R \dot{R} - \dot{S}) \geq 0 . \end{aligned} \tag{2}$$

If, for simplicity, we disregard the last term before the inequality sign, this corresponds to the well-known concept of weak sustainability (Pearce and Atkinson, 1993) according to which the overall capital stock (man-made and natural) should be non-declining. One way to achieve this would entail that *both* K and N should not decline, a requirement commonly referred to as strong sustainability.

The weaker of the two sustainability concepts presupposes that man-made and natural capital can be aggregated since they are to some degree substitutes for each other. As eq. (2) shows, this aggregation requires weights, which are based on the marginal welfare of consumption and nature (W_C and W_N, respectively) and on marginal products. Of course, even for known welfare and production functions, these marginals can probably take almost all non-negative values. In fact, they will depend on the development path actually taken by the economy. The theoretically appropriate marginals will be those along an optimal sustainable path. Unique values can be obtained by solving the intertemporal optimisation problem

$$\max \int_0^T e^{-\rho t} W(Y(K,R) - S, N)\, dt \tag{3}$$

subject to the appropriate equations of motion, typically $\dot{K} = S - \delta K$, $\dot{N} = G(N) - R$, (δ = depreciation rate, $G(N)$ = growth function of natural capital), and the requirements (a) and (b).

The advantage of this approach with respect to the problem of constructing meaningful sustainability indices is that the weights necessary for aggregation are conceptually well-defined and can accommodate any choice of measurement units for K and N. The difficulty, however, is that the weights are hard to determine in

[1] The framework presented here has been chosen for simplicity of exposition. More complex set-ups could be considered.

reality. Not all researchers and policy makers would be prepared to specify and solve problem (3), given the prevailing uncertainties.[2] Whether and to what extent weights determined with standard valuation techniques (such as contingent valuation or hedonic regressions) can be used as surrogates is not to be discussed here.

Constructing a meaningful index may appear easier if strong sustainability is considered instead, in which case natural capital, N, is required to be non-declining. But natural capital is a construct, that is, an aggregate of rather diverse components. In other words, natural capital is itself an index. Conceptually, it could be constructed within the framework sketched above by simple re-interpretation of the symbol N, which would then denote a vector rather than a scalar. The difficulties of operationalising the aggregation, however, would probably be almost as severe as in the case of weak sustainability.[3]

2.2 Bio-Physical Effects-Metric

An important alternative to welfare-based sustainability indices is based on bio-physical cause-effect relationships. In this approach, effects are usually classified according to certain 'themes' or 'issues', which often correspond to categories of environmental damage. Issues frequently considered in the literature are climate change, ozone layer depletion, acidification, eutrophication, toxic contamination, and others (Adriaanse, 1993, OECD 1993). Issue indices can plausibly be categorised as a set of strong sustainability indices, that is, they represent sub-aggregates of natural capital. If these dimensions of natural capital are thought to be complementary to each other, rather than substitutable, the aggregation problem mentioned at the end of the preceding sub-section is not relevant.

An example may serve to illustrate the approach. Consider the issue of eutrophication of water and soil. It is caused by an excessive supply of plant nutrients in the form of phosphates and nitrogen compounds. An index of eutrophication is an aggregate of the loads of the two constituents which cause eutrophication: phosphates, expressed in terms of phosphorus, and nitrates, expressed in terms of nitrogen. The two substances differ with respect to their eutrophication effect: A kiloton (kt) of nitrogen can be taken to have a ten-times smaller effect than a kt of phosphorus (Adriaanse, 1993). Given this effect ratio, either of these substances can be used as a 'numeraire' to measure eutrophication. One possibility is to choose phosphorus as the numeraire and to transform nitrogen loads into 'equivalent' phosphorus loads. A 'eutrophication equivalent' then is 1 kt phosphorus = 10 kt nitrogen, and a eutrophication index can be computed using 1 and 10 as weights. It is trivial how to adjust the weights if one of the substances is measured

[2] Actually, the situation is somewhat paradoxical: Constructing an indicator of sustainability along these lines requires information which the indicator is supposed to deliver, that is, whether the development is sustainable.

[3] A more pragmatic approach to implementing indices of both weak and strong sustainability has been pursued by Pearce and Atkinson (1993).

not in kt but in pounds, say, by invoking the proportionality of the two measurement units.

Given the effect ratios and the possibility to convert measurement units for pollutant loads (kt, kg, pounds), the eutrophication index so defined allows an unambiguous comparison of two (or more) situations with respect to the prevailing degree of eutrophication, independent of measurement units. Needless to say, this is also the case if nitrogen is chosen as the numeraire, rather than phosporus.

The same logic applies to the other environmental issues mentioned above. In the case of climate change, for instance, so-called global warming potentials allow to express the various greenhouse gases in terms of global warming equivalents and to construct indices of global warming pressure. These then permit an unambiguous assessment of whether global warming pressure has increased or decreased.

It can thus be concluded that 'issue indices' are meaningful given that they are based on known scientific relationships which allow to accommodate any legitimate choice of measurement units.

3 Incommensurability: The Problem and Common Approaches[4]

We now consider cases in which the situations to be compared are described neither in terms of welfare nor in terms of well-defined bio-physical damage categories or life support functions. These cases often involve concepts that are inherently vague, such as 'air pollution' or 'water pollution'. In contrast to acidification, eutrophication, and other 'issues' mentioned above, these phenomena are ill-defined and, hence, not directly measurable. This circumstance prevents to derive bio-physical relationships between the constituents that contribute to the respective phenomenon.

In addition to ill-defined phenomena, incommensurability may also arise when it is attempted to aggregate several well-defined 'issues' into an encompassing index. Such an aggregation may be desired especially when a one-dimensional index of (weak or strong) sustainability is sought for.

Both cases have in common that some exogenous weights must be applied to the constituents of the index. Weights may be based, e.g., on opinion polls or expert judgements. Sometimes, explicit weighting is avoided for lack of information, but implicitly this means that equal weights are accorded to the constituent variables. The determination of weights is not the subject of this paper. Weights are thus throughout assumed to be given.[5]

4 Subsections 3.1 and 3.2 are based on Ebert and Welsch (2002). Subsection 3.3 draws on discussions with Udo Ebert.

5 In aggregating theme indices to an encompassing index, weights may differ from country to country. For instance, if some issue is not relevant in a particular country, it may get a weight of zero.

3.1 The Problem

In order to illustrate the problem, consider a simple example which involves the aggregation of two environmental issues. Similar problems may arise when it is attempted to aggregate several polluting substances into an index of an ill-defined phenomenon like air pollution or water pollution.

Suppose we want to assess whether the joint environmental pressure from acidification (sulphur) and eutrophication (phosphorus) has decreased or increased between $t = 1$ and $t = 2$. The data are given in **Table 1**. They are taken from Adriaanse (1993) and refer to the Netherlands 1980 and 1985.

Table 1. Illustrative data for acidification and eutrophication

	Sulphur dioxide in kg per hectare	Phosphorus in kt
t = 1	210	306
t = 2	200	310

For simplicity we assume that the two types of environmental pressure are to be weighted equally. Choosing as an index formula the arithmetic mean yields 258 in $t = 1$ and 255 in $t = 2$. Overall environmental pressure has thus *decreased*, or so it seems. However, we could have chosen different units for one or both types of pollution. Suppose that eutrophication were measured in units of 100 kg of phosphorus, instead of kilotons as assumed above. The phosphorus load would then be 3060 and 3100 in $t = 1$ and $t = 2$, respectively, and the associated index values would be 1635 and 1650, indicating an *increase* of environmental pressure. We cannot thus be sure whether the state of the environment, composed of acidification and eutrophication, has improved or deteriorated.

A similar ambiguity would have occurred had we chosen to measure eutrophication not in phosphorus equivalents, but in nitrogen equivalents. The level of eutrophication would then be 3060 kt nitrogen in $t = 1$ and 3100 kt nitrogen in $t = 2$, and the joint index would again have *increased* from 1635 to 1650. Ambiguity may thus arise not only from trivial choices of units (kilotons, kilograms, pounds etc.) but also from the choice of the substance (the 'numeraire') in which to express some type of environmental pressure.

The question arises whether ambiguity can be avoided by choosing a different index formula (keeping the weighting scheme unaltered). One possibility would be to choose the geometric rather than the arithmetic mean. In this case the index values for $t = 1$ and $t = 2$, respectively, are $(210*306)^{1/2} = 253.5$ and $(200*310)^{1/2} = 249.0$ under the initial choice of units. Under the alternative choice of units, the values $(210*3060)^{1/2} = 801.6$ and $(200*3100)^{1/2} = 787.4$ would be obtained. Independent of measurement units for the individual environmental pressures this kind of index unambiguously indicates a decrease in overall pressure.

This example has shown that some index formulas may give rise to ambiguous comparisons while others allow to avoid these problems. Section 4 will examine systematically which indices are appropriate ('meaningful') in which circumstances.

The problems just illustrated are, to some extent, rather obvious, and there are several ways how they are approached in practice. These can be classified into cardinal and ordinal approaches. The remainder of this section discusses these approaches.

3.2 Cardinal Approaches

The cardinal approach to index construction involves a two-step procedure. It consists of converting the variables from their original (natural) units to 'normalised' (artificial) units and then aggregating the results. The rationale put forward for the first step is that the crude environmental data X_i are considered not suitable for direct aggregation because they may differ with respect to their size (units of measurement) and their variability (range). Obviously, for any given aggregator function and explicit weighting scheme, the *effective* weight of any variable i may depend on the units in which it is measured and on the range it occupies. This may affect the index value in the way illustrated above and imply ambiguous comparisons.

The first step, normalisation, involves in most cases a linear transformation of the crude data, comprising the two elementary operations of translation (addition or subtraction of a constant from all observations of a given variable, thus shifting the origin) and/or expansion (multiplication or division of all values by a constant, thus changing the scale). Following Ott (1978) the normalised variables will be referred to as indicators. The indicator corresponding to the crude data X_i will be denoted by $I_i(X_i)$.

In practice, a variety of normalisation procedures are being applied. Two broad categories can be distinguished: Ranging and standardisation. *Ranging* scales the crude data into the interval 0 to 1. In these approaches, the largest observation has the value 1, but the smallest observation may or may not have the value 0:

$$I_i(X_i) = X_i / X_i^{\max} \quad \text{or} \tag{4}$$

$$I_i(X_i) = (X_i - X_i^{\min}) / (X_i^{\max} - X_i^{\min}) \tag{5}$$

where $X_i^{\min}$ and $X_i^{\max}$ denote the minimum and, respectively, maximum of variable i. The first version of ranging (eq. (4)) retains the origin, i.e. zero is mapped to zero. In the second version (eq. (5)) the smallest observation is mapped to zero.

In *standardisation*, indicator values are obtained by subtracting the mean (μ) from the observations and dividing by the standard deviation (σ):

$$I_i(X_i) = (X_i - \mu_i) / \sigma_i \tag{6}$$

Standardised values, hence, give the deviation of the corresponding underlying variable from the mean of observations, expressed in standard deviation units.

Clearly, the standardised variables are mixed in sign even if the crude observations are all positive.

A common problem of these approaches is that the mean, the standard deviation, the smallest and the largest value can all change as additional observations become available. This renders these indices inappropriate for comparisons over time, which are essential to any sustainability assessment. This problem is avoided if, instead of parameters of the observed distribution, exogenous reference values or standards are employed in normalisation. For instance, the $X_i^{\min}$ and $X_i^{\max}$ in eq. (4) or (5) could be replaced by certain fractions or multiples of environmental standards (or target values). However, even though such indicators and the resulting indices are not plagued with the observation-dependence of the normalisation parameters, the choice of exogenous normalisation parameters is never free of arbitrariness.

Arbitrariness also widely prevails with respect to the aggregation rule applied in the second step. Some authors justify their choices by considerations of presumed substitutability among the environmental variables (see Ott, 1978, Khanna 2000). Widely used aggregation rules are the arithmetic mean, the geometric mean, and the constant-elasticity-of-substitution function (of which the former two are limiting cases). In practice, these aggregation rules are combined with diverse forms of normalisation in a largely unsystematic way, as the examples in **Table 2** illustrate.

Table 2. A selection of environmental indices

Reference	Normalisation	Aggregation
Adriaanse (1993)	$I_i(X_i) = X_i / X_i^*$	$I(X) = \Sigma_i I_i(X_i)$
Den Butter/van der Eyden (1998)	$I_i(X_i) = X_i / X_i^{1980}$	$I(X) = \Sigma_i w_i I_i(X_i)$
ESI (2001)	$I_i(X_i) = (X_i - \mu)/\sigma_i$	$I(X) = \frac{1}{n}\Sigma_i I_i(X_i)$
Hope et al. (1992)	$I_i(X_i) = X_i / X_i^{1980}$	$I(X) = \Sigma_i w_i I_i(X_i)$
Khanna (2000)	$I_i(X_i) = \dfrac{(X_i - \bar{X}_i^{\min})}{(\bar{X}_i^{\max} - \bar{X}_i^{\min})}$	$I(X) = \left(\frac{1}{n}\Sigma_i I_i(X_i)^{\varepsilon}\right)^{1/\varepsilon}$
Van der Bergh/van Veen-Groot (2001)	$I_i(X_i) = X_i / X_i^{\max}$	$I(X) = \frac{1}{n}\Sigma_i I_i(X_i)$

Adriaanse (1993) normalises the values of his 'theme' indices (climate change, ozone layer depletion, acidification etc.) by dividing them by a target value, and adds the normalised values across themes. The pollution index of Hope et al. (1992) uses as its input variables pollutant loads in water, air, and soil, which are expressed as a fraction(or multiple) of their respective values in a base year. These normalised values are then aggregated using the weighted arithmetic mean, with weights derived from opinion polls. Similar normalisations are applied by den

Butter and van der Eyden (1998) and van den Bergh and van Veen-Groot (2001). The 'Environmental Sustainability Index' of the World Economic Forum (ESI, 2001) employs standardisation as a normalisation device. They all use the arithmetic mean to aggregate their normalised data. Khanna (2000) chooses a constant-elasticity-of-substitution aggregator function.

Note that in the case of Adriaanse (1993) and den Butter/van der Eyden (1998) X_i denotes the value of a 'theme indicator' in 'theme equivalent units' and $I_i(X_i)$ the normalised theme indicator value. In all other cases the X_i are in natural units. X_i^* denotes an exogenous target value, and $\bar{X}_i^{min}$ and $\bar{X}_i^{max}$ are 50% and 500% of the National Ambient Air Quality Standards in the U.S. w_i denotes the weight of variable i.

It is obvious that any two of the normalisation-aggregation configurations in **Table 1** may imply different comparisons of states of the environment since these procedures approach the problem of measurement in rather different, yet arbitrary ways.

3.3 Ordinal Approaches

The ordinal approach deals with the problems implied by the non-uniqueness of the cardinalisation of variables by simply ignoring the cardinal dimension of the data and focusing on the ranking of the objects to be compared.[6] In the case of sustainability assessment, the 'objects' are the environmental situations at $t=1$ and $t=2$, which are described in terms of at least two variables (or attributes).

The ordinal approach consists in aggregating the rank orders of the objects by individual attributes, instead of aggregating the numerical values of the attributes by individual object. A variety of rank aggregation procedures have been proposed (see Chebotarev and Shamis 1998 for an overview and characterisation). They all share a common drawback, namely that they are subject to the impossibility theorem of Arrow (1963) which states that there exists no aggregation rule for rank order preferences that satisfies a set of reasonable axioms. However, one of the Arrow axioms is of little importance in many applications of ranking rules: the 'independence of irrelevant alternatives'. To see this, consider that the 'alternatives' correspond to the objects to be ranked. Whenever the set of objects is fixed (e.g. a fixed list of countries for which a rank order is being sought), the axiom is itself rather 'irrelevant' unless the set of objects is to be changed.[7]

Unfortunately, the axiom *cannot* be discarded if one attempts to apply ranking rules in sustainability assessment. In fact, ranking rules may imply inconsistent assessments. This can be illustrated for the case of the Borda rule, which is the probably simplest and best known ranking rule.

---

6 In fact, ordinal approaches can be applied to data which are ordinal in character, that is, data which lack any cardinal significance.

7 Examples of country rankings include Dasgupta and Weale (1992) and Paul (1997).

The Borda rule is based on simple addition across attributes of the attribute-specific ranks of the objects.[8] More specifically, the Borda score of an object is given by

$$B = \Sigma_k (n - a_k) \tag{7}$$

where n is the number of objects and a_k is the rank number of the considered object in respect of attribute k (higher value of attribute is mapped to higher a_k). The Borda score is 1 for the highest-ranking object and n for the lowest-ranking one.

If we wish to employ the Borda score for sustainability assessment, the objects correspond to time periods t_i and t_j, and $n = 2$. However, there is a sequence of such bilateral comparisons: t_1 vs. t_2, t_2 vs. t_3, etc. Consistency would, of course, require that an improvement in t_2 relative to t_1 and an improvement in t_3 relative to t_2 implies an improvement in t_3 relative to t_1 (transitivity). However, since the 'independence of irrelevant alternatives' is not satisfied, transitivity is not guaranteed.[9]

A numerical example may illustrate this. Assume t_1, t_2, and t_3 are to be compared with respect to 3 pollutants X_1, X_2, and X_3. The (hypothetical) data are given in **Table 3**. The measurement units are ignored since we want to apply an ordinal comparison method.

Table 3. Illustrative pollution data

	X_1	X_2	X_3
t_1	2	6	3
t_1	4	4	4
t_3	1	5	5

At t_2 we can compare t_2 with t_1. The respective Borda scores are $B(t_1) = 2$, $B(t_2) = 1$, that is t_2 ranks first in terms of overall environmental pressure. In other words, environmental pressure has *increased* between t_1 and t_2. As time proceeds, we can compare t_3 with t_2, which yields $B(t_2) = 2$, $B(t_3) = 1$. Thus, there is a *further increase* in environmental pressure. However, a *direct* comparison of t_3 and t_1 yields $B(t_1) = 1$ and $B(t_3) = 2$, indicating a *decline* of pollution.

The Borda ranking rule thus fails to produce unambiguous comparisons of environmental pressure over time. The same is true for all ordinal methods of comparison, the reason being that they violate the axiom of 'independence of irrelevant alternatives'.

8 Of course, it would be possible to apply weights to the individual attributes before adding them.

9 This is akin to the Condorcet paradox in social choice theory.

4 Incommensurability: The Social Choice Approach

It may have become clear that common approaches to the construction of sustainability indices involving incommensurable variables all imply the risk of yielding ambiguous (contradictory) assessments. In this sense, these indices are not meaningful. I will now address in a systematic way the question, which indices are meaningful in which circumstances, given that the input variables are incommensurable.[10]

4.1 Basic Framework

We consider $n \geq 2$ environmental variables $X_1,...,X_n$ measuring environmental quality at some point or period of time. Let a vector $X = (X_1,...,X_n)$ denote a quality profile or environmental state. Moreover, we consider a preference ordering *P*. The ordering is defined for all states in the admitted domain $X \in D^n$. It reflects a researcher's or social decision maker's value judgements on states of the environment.

We suppose that the ordering can be represented by an environmental index *I*: $D^n \rightarrow \Re$; i.e. we have

$$X^1 \;\; P \;\; X^2 \Leftrightarrow I(X^1) \geq I(X^2) \text{ for all } X^1, X^2 \in D^n. \tag{8}$$

The index is only ordinally unique, i.e., every strictly increasing transformation of *I*(*X*) is also a representation of the preference ordering. This is, of course, sufficient for a sustainability index since we are only interested in whether the environmental situation has improved or deteriorated over time.

As stated above, ambiguous (contradictory) comparisons of environmental states may arise due to the non-uniqueness of measurement scales. Therefore one should take into consideration the possibility of changing the scales by which the variables X_i can be measured.

Changing scales comes down to transforming the quality profile by a corresponding transformation $\Phi = (f_1,...,f_n)$

$$\Phi : (X_1,...,X_n) \rightarrow (f_1(X_1),...,f_n(X_n)) \tag{9}$$

where f_i, $i = 1,...,n$, reflects the respective operation. Let $F := \{\Phi = (f_1,...,f_n) | f_i : D \rightarrow D \text{ for } i = 1,...,n\}$ be the set of transformations of D^n into itself which are admitted.

We request that the ranking of environmental profiles must not depend on the choice of scales, i.e. must not be changed by any admissible transformation:

10 This section is based on Ebert and Welsch (2004), to which the reader is referred for details and further references.

$$X^1 \ P \ X^2 \Leftrightarrow \Phi(X^1) \ P \ \Phi(X^2) \tag{10}$$

for all $X^1, X^2 \in D^n$ and $\Phi \in F$. In other words, the ranking has to be invariant with respect to admissible transformations. If an ordering P satisfies (3) it is called invariant with respect to F (or F-invariant).

The invariance condition provides the basis of our definition of a meaningful index, since it can equivalently be stated in terms of any representation I of P:

$$I(X^1) \geq I(X^2) \Leftrightarrow I(\Phi(X^1)) \geq I(\Phi(X^2)) \tag{11}$$

for all $X^1, X^2 \in D^n$ and $\Phi \in F$.

We call an index *meaningful* if it satisfies this condition: An environmental index is meaningful if the underlying preference ordering is invariant with respect to admissible transformations of the environmental variables.[11]

Below we will provide characterisations of classes of indices which are meaningful given that the environmental variables possess certain measurability and comparability properties. To that purpose we will now classify environmental variables into categories of measurability and comparability.

4.2 Measurability and Comparability

Measurability relates to the question what kinds of transformations are admissible for any *single* variable, whereas comparability is concerned with the question whether *several* variables can be transformed independently.

With respect to measurability the classes of interest are interval-scale measurable and ratio-scale measurable variables. A variable X_i is interval-scale measurable if its ordering is unique up to a transformation of the form $f_i(X_i) = \alpha_i X_i + \beta_i$ for $\alpha_i > 0$. In other words, admissible transformations involve translations as well as expansions. Interval-scale measurable variables allow to compare levels and differences independent of transformations of the form stated above, but do not allow to compare ratios. This is different with ratio-scale measurable variables:

A variable X_i is ratio-scale measurable if its ordering is unique up to a transformation of the form $f_i(X_i) = \alpha_i X_i$ for $\alpha_i > 0$. Hence, only expansions but not translations are admissible, that is, the variable has a fixed (natural) origin. The consequence is that levels, differences, and ratios can be compared, independent of transformations of the form stated above.

With respect to comparability, five cases may be relevant for sustainability assessment.

11 The property of 'meaningfulness' is a purely technical one, not suggesting any substantive connotation. The term comes from 'measurement theory', see Pfanzagl (1971), Roberts (1979) or Luce et al. (1990).

(a) Interval-scale noncomparability INC: $f_i(X_i) = \alpha_i X_i + \beta_i$, $\alpha_i > 0$. Under INC, several variables can be translated and expanded separately (independently).

(b) Interval-scale unit comparability IUC: $f_i(X_i) = \alpha X_i + \beta_i$, $\alpha > 0$. Under IUC, several variables can be translated independently, but they can only be expanded jointly.

(c) Interval-scale full comparability IFC: $f_i(X_i) = \alpha X_i + \beta$, $\alpha > 0$. Under IFC, only common translations and expansions are admitted.

(d) Ratio-scale noncomparability RNC: $f_i(X_i) = \alpha_i X_i$, $\alpha_i > 0$. Under RNC, several variables can be expanded independently; translations are not admitted.

(e) Ratio-scale full comparability RFC: $f_i(X_i) = \alpha X_i$, $\alpha > 0$. Under RFC, several variables can only be expanded jointly.

Among these cases, INC and RNC are perhaps the most important ones. INC applies to temperatures: Temperatures in several situations can be measured in Centigrades, Fahrenheit, Réaumur, and Kelvin, which are related to each other as stated under (a). RNC applies to masses (such as pollutant loads): Masses in several situations can be expressed in pounds, kilograms, tons etc., which are related to each other as stated under (d).

4.3 Results

The results for interval-scale measurability and ratio-scale measurability are presented in **Table 4** and **Table 5**, respectively.

Table 4. Results for interval-scale measurability

	C+WM	C+SM	C+SM+SEP
INC	dictatorial ordering	no ordering	
IUC	$\sum_{i=1}^{n} w_i X_i,\ w_i \geq 0$	$\sum_{i=1}^{n} w_i X_i,\ w_i > 0$	
IFC	$\sum_{i=1}^{n} w_i X_i,\ w_i \geq 0$ and more complicated forms	$\sum_{i=1}^{n} w_i X_i,\ w_i > 0$ and more complicated forms	$\sum_{i=1}^{n} w_i X_i,\ w_i > 0$

Note: C, WM, SM, and SEP denote continuity, weak monotonicity, strong monotonicity, and separability. These are properties that one may wish to impose on the preference ordering.

It can be seen from **Table 4** that for interval-scale non-comparable (INC) variables at best a dictatorial ordering (dictatorial index) exists, depending on the desired type of monotonicity. The frequently used arithmetic mean requires at least interval-scale unit comparability, that is, the scale of the variables can only be changed jointly (by a common factor); only the origin may be shifted independently.

For ratio-scale measurable variables, the prospects for obtaining a meaningful index are more favourable (see **Table 5**). Most interesting is the case of ratio-scale non-comparable variables, like pollutant loads. In this case, the geometric mean (Cobb-Douglas function) provides a meaningful index

Table 5. Results for ratio-scale measurability

	C+WM	C+SM	C+SM+SEP
RNC	$\prod_{i=1}^{n} X_i^{w_i}, w_i \geq 0$ on $\Re_+^n$ or $\Re_{++}^n$	$\prod_{i=1}^{n} X^{w_i}, w_i > 0$ on $\Re_{++}^n$	
RFC	any homothetic function	any homothetic function	CES function

Note: See **Table 4**.

In the case of mixed measurability, chances to obtain a meaningful index are poor. For instance, if one wishes to define an ordering over an INC variable and a RNC variable the ordering must be dictatorial to be meaningful (free of ambiguity). This then implies that either the INC variable or the RNC variable must be chosen as an index. It is impossible to *combine* them in a meaningful way. This case is, e.g., relevant for water quality indices, which often aim at combining temperature (INC) with pollutant loads (RNC).

5 Conclusions

The paper has surveyed and evaluated the possibilities and limitations of sustainability indices from the point of view of meaningfulness. A meaningful sustainability index is one which allows unambiguous orderings of the relevant 'situations' over time, independent of the measurement units in which the variables describing the situations are expressed. The cases of commensurability and incommensurability were distinguished. In the former, the comparison of situations is unambiguous because all legitimate choices of measurement units can be accommodated on the basis of exogenously given relationships among the variables. These relationships may define a welfare-based monetary metric or an effects-based bio-physical metric. In the case of incommensurability, common approaches (both cardinal and ordinal) may fail to yield meaningful indices.

A systematic assessment of which indices are meaningful in which circumstances has shown that indices in the frequently used form of an arithmetic mean are meaningful only if the variables satisfy (at least) interval-scale unit comparability, i.e. they can be scaled only by a common factor. Important environmental variables like emissions or pollutant loads (masses) do not possess this property. Masses are ratio-scale non-comparable, that is, they can be scaled independently (pounds, kilograms, tons), but have a fixed origin (0 pounds = 0 kilograms etc.). An environmental index which combines masses using the geometric mean is meaningful. Masses (RNC) and temperatures (INC) cannot be combined with each other – as attempted in some water quality indices – in a meaningful way.

The criterion that sustainability indices should guarantee unambiguous comparisons of situations is intuitively plausible. Yet, there may exist a tension with other criteria. For instance, the geometric mean (Cobb-Douglas function) implies that the elasticity of substitution between the constituent variables (e.g. pollutant loads) is unity. This implied elasticity may differ from *a-priori* ideas a researcher or policy maker might have regarding substitutability. A problem may then arise especially when the substitutability that is deemed appropriate is lower than unity. However, the frequently applied arithmetic mean is even worse in this regard: It implies an infinite elasticity of substitution, that is, any deterioration with respect to one pollutant can be neutralised by a finite improvement with respect to another pollutant. In comparison with this case the geometric mean provides a relatively 'prudent' type of index.

If the conditions for meaningfulness are not fulfilled, all one can probably do is to take resort to a statistical approach: One could compute a battery of indices which involve a variety of normalisations and aggregation rules and then show the distribution of improvements vs. deteriorations. This would yield a sort of likelihood that the 'situation' has improved or deteriorated. Possibly, a unanimous answer may be obtained, and hopefully it will indicate that the requirement of sustainability is satisfied.

References

Adriaanse, A. (1993), *Environmental Policy Performance Indicators*, SDU, Den Haag.

Arrow, K.J. (1963), *Social Choice and Individual Values*, New York.

Chebotarev, P.Y. and E. Shamis (1998), Characterizations of Scoring Methods for Preference Aggregation, *Annals of Operations Research* 80, 299-332.

Dasgupta, P.S. and M. Weale (1992), On Measuring the Quality of Life, *World Development* 20, 119-131.

Den Butter, F.A.G. and J.A.C. van der Eyden (1998), A Pilot Index for Environmental Policy in the Netherlands, *Energy Policy* 26, 95-101.

Ebert, U. and H. Welsch (2004), Meaningful Environmental Indices: A Social Choice Approach, *Journal of Environmental Economics and Management* 47, 270-283.

Ebert, U. and H. Welsch (2002), *Meaningful Environmental Indices: A Social Choice Approach*, Discussion Paper V-242-02, Department of Economics, University of Oldenburg.

ESI (2001), *Environmental Sustainability Index 2001*, An Initiative of the Global Leaders of Tomorrow Environment Task Force, www.ciesin.columbia.edu/indicators/ESI.

Hope, C., J. Parker, and S. Peake (1992), A Pilot Environmental Index for the UK in the 1980s, *Energy Policy* 20, 335-343.

Khanna, N. (2000), Measuring Environmental Quality: An Index of Pollution, *Ecological Economics* 35, 191-202.

OECD (1993), *OECD Core Set of Indicators for Environmental Performance Reviews: A Synthesis Report by the Group on the State of the Environment*, Environment Monographs No. 83, Paris.

Luce, R.D., D.H. Krantz, P. Suppes, and A. Tversky (1990), *Foundations of Measurement,* Vol. III, San Diego.

Ott, W. (1978), *Environmental Indices: Theory and Practice*, Ann Arbor.

Pfanzagl, J. (1971), *Theory of Measurement*, 2nd edition, Würzburg.

Roberts, F.S. (1979), *Measurement Theory with Applications to Decision-Making*, Utility and the Social Sciences, Reading.

Paul, S. (1997), The Quality of Life: An International Comparison Based on Ordinal Measures, *Applied Economics Letters* 4, 411-414.

Pearce, D. and G. Atkinson (1993), Capital Theory and the Measurement of Sustainable Development: An Indicator of Weak Sustainability, *Ecological Economics* 8, 103-108.

Van den Bergh, J.C.J.M. and D.B. van Veen-Groot (2001), Constructing Aggregate Environmental-Economic Indicators: A Comparison of 12 OECD Countries, *Environmental Economics and Policy Studies* 4, 1-16.