

Fondazione Eni Enrico Mattei

**On the Art of Aggregating  
Apples & Oranges**

Jochen Jesinghaus

**NOTA DI LAVORO 91.2000**

Corso Magenta, 63, 20123 Milano, tel. +39/02/52036934 – fax +39/02/52036946

E-mail: [letter@feem.it](mailto:letter@feem.it)

C.F. 97080600154

# On the Art of Aggregating Apples & Oranges

Prepared for the EVE Workshop, March 6 & 7, 2000 at Fondazione Eni Enrico Mattei (FEEM, Milano) by Jochen Jesinghaus, European Commission JRC/ISIS, I-21020 Ispra (VA)

The following paper reflects the author's personal thoughts, which are not necessarily identical with official positions of the European Commission. Comments to [Jochen.Jesinghaus@jrc.it](mailto:Jochen.Jesinghaus@jrc.it)

<b>1. Goal: better decisions</b>	<b>3</b>
1.1 Analytical framework: The Driving forces-Pressure-State-Impact-Response model	3
1.2 Case study: The European Environmental Pressure Indices project	4
<b>2. How do indicators influence decision-making?</b>	<b>5</b>
2.1 The Power of Economic Indicators	5
2.2 How indicators influence decisions of citizens	6
2.3 How indicators influence decisions of policy-makers	7
<b>3. Better than GDP: a Policy Performance Index (PPI)</b>	<b>10</b>
3.1 Are we allowed to aggregate “apples and oranges”?	10
3.2 Policy Performance Indices encourage politicians to make good decisions	10
3.3 Defining the <b>share</b> of the PPI components	11
3.4 Defining the <b>valuation</b> of the PPI components	13
<b>4. Costs &amp; Benefits of Sustainable Development Realpolitik</b>	<b>16</b>
4.1 How much are you Willing To Sacrifice for a green light?	18
4.2 How much does it Cost to Avoid losing the elections?	18
4.3 Last but not least: why is there no science in the picture?	19
<b>5. Annex: Basic Guide for Constructing a Policy Performance Index</b>	<b>19</b>
5.1 How many issues?	19
5.2 Which issue sets?	19
5.3 How many indicators per issue?	19
5.4 Which indicators?	19
5.5 Which weights?	20
5.6 How to value a trend as “policy success” or “policy failure”?	20

## List of figures:

Figure 1: The Driving force-Pressure-State-Impact-Response model	3
Figure 2: The indicator set of the European Environmental Pressure Indices project	4
Figure 3: The indicator-voters interface	6
Figure 4: Indicators, media, voters and politics	8
Figure 5: “Reduced complexity” and today’s information system	9
Figure 6: Defining the shares of the PPI’s environmental sub-index	12
Figure 7: Inside an Environmental Pressure Index: an Air Pollution Sub-Index	13
Figure 8: Relative valuation against past policy performance: the unemployment example	14
Figure 9: Comparison to countries of the same class: the CO <sub>2</sub> emission reductions example	15
Figure 10: Ecosistema Urbano: comparison of Italian cities with each other	15
Figure 11: The political dimension of damage and avoidance costs	17
Figure 12: Asking the right questions: WTP for what? Avoiding what?	17
Figure 13: The average citizen's WTP for the environment	18
Figure 14: Avoidance costs: the CO <sub>2</sub> example	18

# 1. Goal: better decisions

About five years ago, the *Communication from the Commission to the Council and the European Parliament on "Directions for the EU on Environmental Indicators and Green National Accounting"* (COM (94) 670 final, 21.12.94) laid the basis for an ambitious project aimed to give comprehensive support to environmental policy and consisting of the following “key actions”:

- Developing a European System of Environmental Pressure Indices;
- Developing Integrated Economic and Environmental Indices;
- Developing Environmental Satellite Accounts;
- Research on damage evaluation and monetization techniques.

This paper will examine the question how the indicator-related “key actions” can be better linked to the research on monetary evaluation, and in particular the ongoing efforts by the GARP (damage estimation using Impact Pathway Analysis and WTP) and GreenStamp (monetisation through Avoidance Costs and economic modelling) projects.

The author will try to demonstrate:

- that indicators are a powerful driving force of many, if not most, political decisions;
- that bad indicators are thus a recipe for bad politics;
- how an indicator system that serves democratic decision-making should be designed;
- finally, how GARP & GreenStamp could be applied to such an indicator system.

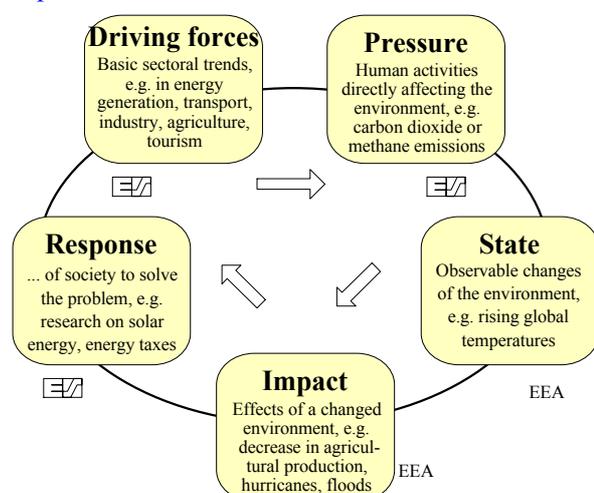
In particular, it will try to demonstrate that “Green Accounting” can become the bridge between citizens and decision-makers, if applied in a transparent and policy-relevant way (this does not categorically exclude monetisation, but “transparency” is an ambitious goal for monetary valuation projects).

## 1.1 Analytical framework: The Driving forces-Pressure-State-Impact-Response model

Environmental and Sustainable Development policy is a fairly complex policy area and therefore needs both an appropriate analytical framework, and a comprehensive and detailed set of indicators. The most widely accepted indicator framework is the “Driving forces-Pressure-State-Impact-Response model”<sup>1</sup>, which defines five indicator categories:

Figure 1: The Driving force-Pressure-State-Impact-Response model

- D Driving forces** are underlying factors influencing a variety of relevant variables. Examples: the number of cars per inhabitant; total industrial production; GDP.
- P Pressure** indicators describe the variables which directly cause environmental problems. Examples: toxic emissions, CO<sub>2</sub> emissions, noise etc. caused by road traffic; the parking space required by cars; the amount of waste produced by scrap cars.
- S State** indicators show the current condition of the environment. Examples: the concentration of lead in urban areas; the noise levels near main



<sup>1</sup> The DPSIR model is an extension of the PSR (Pressure-State-Response) model, developed by Anthony Friend in the 1970s, and subsequently adopted by the OECD’s State of the Environment (SOE) group.

roads; the global mean temperature.

I **Impact** indicators describe the ultimate effects of changes of state. Example: the percentage of children suffering from lead-induced health problems; the mortality due to noise-induced heart attacks; the number of people starving due to climate-change induced crop losses.

R **Response** indicators demonstrate the efforts of society (i.e. politicians, decision-makers) to solve the problems. Examples: the percentage of cars with catalytic converters; maximum allowed noise levels for cars; the price level of gasoline; the revenue coming from pollution levies; the budget spent for solar energy research.

## 1.2 Case study: The European Environmental Pressure Indices project

Within the DPSIR framework, Eurostat (the Statistical Office of the European Communities) focuses on the *Driving forces*, *Pressure* and *Response* categories <sup>2</sup>.

The Environmental Pressure Indices Project, conducted by Eurostat and financed by the European Commission’s Environment DG, aims at a comprehensive description of the most important human activities that have a negative impact on the environment. The project reflects the efforts undertaken by the European Commission to provide decision-makers and the general public with the information necessary for the design and monitoring of an adequate environment policy for the European Union. The first indicator publication, “Towards Environmental Pressure Indicators” (Eurostat 1999), covers the following 60 indicators: <sup>3</sup>

Figure 2: The indicator set of the European Environmental Pressure Indices project

Air Pollution	Emissions of nitrogen oxides (NOx)	Emissions of non-methane volatile organic..	Emissions of sulphur dioxide (SO2)	Emissions of particles	Consumption of gasoline & diesel oil by road vehicles	Primary energy consumption
Climate Change	Emissions of carbon dioxide (CO2)	Emissions of methane (CH4)	Emissions of nitrous oxide (N2O)	Emissions of chlorofluorocarbons (CFCs)	Emissions of nitrogen oxides (NOx)	Emissions of sulphur oxides (SOx)
Loss of Biodiversity	Protected area loss, damage and fragmentation	Wetland loss through drainage	Agriculture intensity: area used for intensive..	Fragmentation of forests & landscapes..	Clearance of natural & semi-natural forested..	Change in traditional land-use practice
Marine Environment & Coastal Zones	Eutrophication	Fishing pressure	Development along shore	Priority habitat loss	Discharges of heavy metals	Oil pollution at coast & at sea
Ozone Layer Depletion	Emissions of bromofluorocarbons (halons)	Emissions of chlorofluorocarbons (CFCs)	Emissions of hydrochlorofluorocarbons..	Emissions of carbon dioxide (CO2)	Emissions of nitrogen oxides (NOx)	Emissions of chlorinated carbons
Resource Depletion	Water consumption per capita (incl. ground..	Use of energy per capita	Increase in territory permanently occupied by..	Nutrient balance of the soil (nutrient..	Electricity production from fossil fuels (mineral..	Timber balance (new growth/..
Dispersion of Toxic Substances	Consumption of pesticides by agriculture	Emissions of persistent organic pollutants..	Consumption of toxic chemicals	Index of heavy metal emissions to water	Index of heavy metal emissions to air	Emissions of radioactive material
Urban Environmental Problems	Energy consumption	Non-recycled municipal waste	Non-treated wastewater	Share of private car transport	People endangered by noise emissions	Land use (change from natural to built-up area)
Waste	Waste landfilled	Waste incinerated	Hazardous waste	Municipal waste	Waste per product during a number of..	Waste recycled/material recovered
Water Pollution & Water Resources	Nutrient (nitrogen & phosphorus - N + P) use..	Ground water abstraction	Pesticides used per hectare of agriculture..	Water treated/water collected	Index of heavy metals emissions	Emissions of organic matter as biochemical..

Environmental decision-makers (e.g. officials in environment ministries) who are specialised in one of the ten policy fields (which are basically “clusters” of similar impacts) will recognize each of the six pressure indicators as (often imperfect) representations of important contributions to the “overall problem pressure” ; and they would probably insist that all six indicators, if not more, are necessary to provide an information tool that reasonably covers the contents of their daily work.

<sup>2</sup> As indicated in the graph, the European Environment Agency (EEA) has the lead in the “State” and “Impact” categories. In practice, both organisations cooperate closely, and there is a considerable but inevitable overlap.

<sup>3</sup> These indicators have been identified through mailed surveys among a panel of 2,300 European environmental experts, the so-called Scientific Advisory Groups (SAG).

## 2. How do indicators influence decision-making?

One might conclude that with ...

- a consistent analytical framework like the DPSIR model,
- and a solid, detailed and policy-relevant set of indicators,

the decision-maker (e.g. in an environmental ministry or an environmental agency) has everything that is needed to start a successful working day.

Unfortunately, experience shows that *informed* political decisions are not necessarily *good* decisions. Even a perfect indicator system is not a guarantee that suddenly all the errors that our societies have committed in the past could be avoided. The following examples may illustrate why powerful and well-informed decision-makers are often forced, against their own will, to take wrong decisions.

### 2.1 The Power of Economic Indicators

#### 2.1.1 GDP and economic growth

It is difficult to prove empirically how GDP affects policy decisions; no politician would declare in public “*I had to take the decision X because I wanted to be re-elected, and I knew my voters would judge me on the basis of the GDP growth rate*”. And yet, it is almost impossible to find a newspaper that does **not** contain at least one article lamenting the critical economic situation, and urging the government to take measures to accelerate economic growth, such as lowering the tax burden.

Certainly, the wealth of a nation depends to a great extent on its economic output (and that is basically what GDP measures: *output* valued at market prices); but there is also a broad scientific consensus that GDP should not be misused as a measure of “welfare” or “well-being” of a society. For example, repairing a car after an accident would raise GDP - and car owners would probably not see that as an increase of their personal welfare.

There is abundant literature on why GDP is a flawed measure of economic success, especially with regard to the non-inclusion of external effects. In spite of this criticism, it continues to be the most important “policy performance barometer”: A government may do plenty of good and intelligent things to increase the true welfare of its citizens; but if the GDP growth rate is “minus 2%”, it will definitely lose the next elections. Accordingly, governments try to keep the economy on a growth path, even if that would mean sacrificing other important goals of society.

#### 2.1.2 Inflation rate and carbon tax

Some months ago, the inflation rate in Italy jumped over the magic 2% hurdle, indicating that the Italian economy was no longer following the good example of other leading EU countries. Newspapers and TV stations considered this as a major policy disaster, and soon identified the guilty ones: fuel prices had pushed the inflation rate, as a reaction to OPEC price increases.

The experts also quickly found a solution, and proposed to abolish the long-awaited “carbon tax” on fuel prices. After several weeks of media attacks, the government partly ceased to public pressure, postponed the carbon tax introduction and watered it down to “harmless” rates.

### 2.1.3 Unemployment rate

Rumours say that the government of a major EU Member State that faced elections in September 1998 spent billions of Euros on special measures giving jobs to unemployed people - badly paid jobs, and only for one year or less, but enough jobs to turn the threatening growth of the unemployment rate into a miraculous decrease a few weeks before the elections<sup>4</sup>.

Analogous to the GDP and inflation rate examples, it seems that *any decision that decreases the unemployment rate is a "good" decision* – even if that decision only provides short term jobs (and is thus a waste of taxpayers' money, given that it could be invested in more efficient measures against unemployment like education and training).

#### Conclusions:

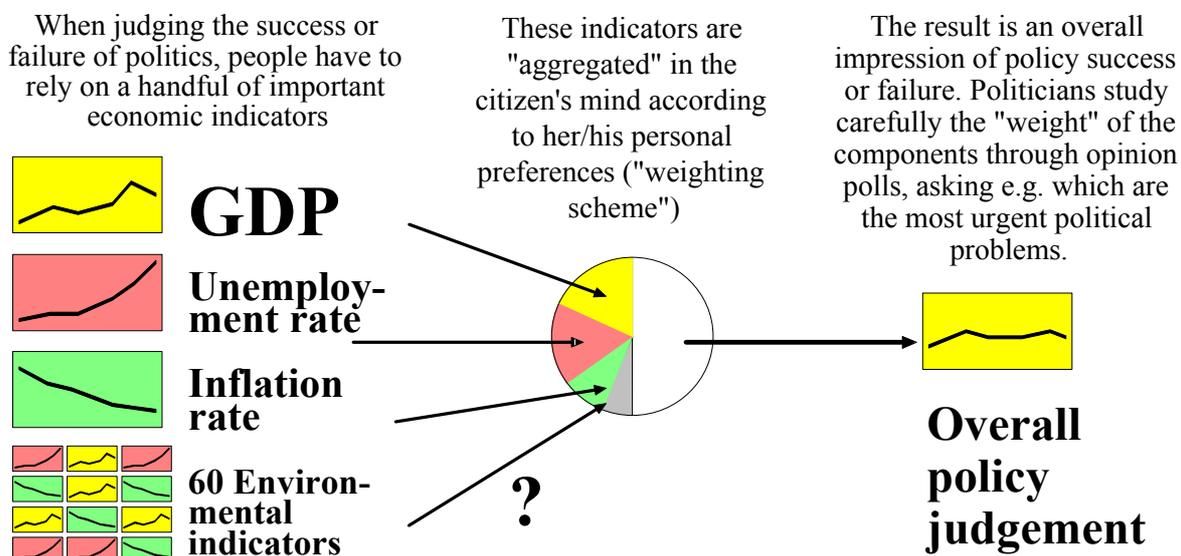
- The “big three” economic indicators (GDP, unemployment rate, inflation rate) have an overwhelming influence on the decision-making of voters.
- A politician who wants to be re-elected **must** improve these indicators.

A politician might start with good intentions and lots of bright ideas how to increase a nation's overall welfare; but four years later, the government wants to be re-elected, and then only the visible, measurable performance counts. Therefore, *any decision that increases GDP, lowers inflation or reduces unemployment is a "good" decision* – even if that decision destroys the environment or increases the gap between the rich and the poor.

### 2.2 How indicators influence decisions of citizens

The following figure illustrates how voters form their opinion on the current government on the basis of policy performance indicators like GDP, unemployment and inflation rates:

Figure 3: The indicator-voters interface



Given that voters do not want to indulge in too much detail, one important observation is that **environmental** policy as a whole will probably not influence the voter's judgement of government performance. In principle, the sixty indicators now produced by Eurostat could introduce the environment into the voters' decision-making process. However, it is obvious

<sup>4</sup> According to Frankfurter Allgemeine Zeitung of 14.08.1998, p. 15, "Institut spricht von Wende am Arbeitsmarkt", such measures had been doubled in the last four months. In Eastern Germany alone, about 138000 additional short-term jobs had been created (with costs probably over 1 Billion Euro). Not enough to convince the voter, however: the elections were won by the social-democratic and green opposition parties.

that a “battery” of 60 environmental pressure indicators, some pointing upwards, others downwards, will confuse non-experts, and will therefore not be taken as seriously as e.g. GDP. Without aggregation to a handful of indices, or even one overall Pressure **Index**, environmental indicators cannot compete with the three well-established economic indicators.

### 2.3 How indicators influence decisions of policy-makers

Certainly, the decision of the voter will not be taken on the basis of indicators alone. Many other factors play a role, like sympathy for a certain candidate, party identification (“*my father was a communist, so am I*”), a perceived need for change or continuity, more or less convincing electoral spots etc.

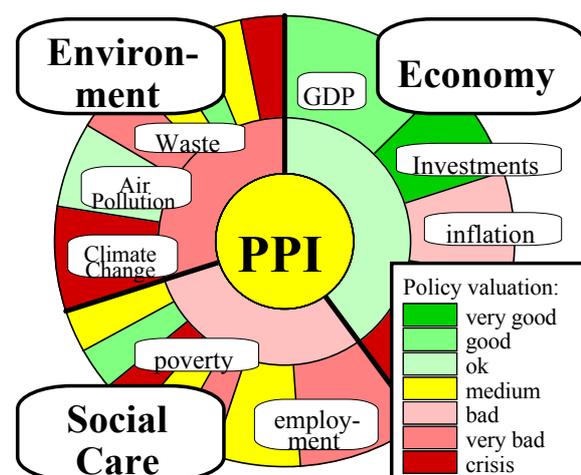
What makes indicators unique as determinants of the voting decision, however, is that they reveal the success or failure of government **decisions**. The charisma of the conservative candidate or the socialist tradition of the voter’s family influence the citizen’s preference independently of the governments real performance. A party may win the elections because the top candidate is particularly good-looking; but there is no causal link between the looks of the candidate and the policy decisions she or he will take. Indicators instead *punish* bad decisions, and *reward* good ones - always with the caveat that the definition of “good” or “bad” decisions comes along with the indicator itself: if the media misuse GDP as a measure of success, then any decision that increases GDP is “good”, even if it decreases real welfare...

In the following section we will examine in more detail how indicators, through their influence on the decision of the voter, make pressure on the decision-making process of governments and parties. We will start with a look in the future, and assume that the following indicators are available and widely used by the media:

- one **Environmental** Pressure Index, sub-divided in ten “policy field indices”, each of which composed of six indicators (i.e. a total of 60 components).
- a similar index covering **social** issues, like the quality of health services, income distribution and poverty, education etc.;
- an analogous index for “**economic** performance” consisting of typical indicators such as GDP, inflation and investment rates.

All three indices are aggregated to a “Policy Performance Index”, PPI, and presented as a pie chart organised in three concentric circles as follows:

- the three levels contain: 1) one overall index (PPI), 2) three sub-indices for “Economy”, “Social Care” and “Environment” and 3) an outer circle representing sub-sub-indices or “simple” indicators such as GDP, inflation, poverty rate <sup>5</sup>, Climate Change, Waste and Air Pollution pressure index;
- the size of each segment reflects the importance (the “weight”) of this issue for politics;
- the colour of each segment reflects the judgement of current policy performance on a seven colour scale, i.e. green for “good” and red for “bad”. The inner two levels are aggregated valuations of the underlying segments (i.e. the “yellow”= “medium” PPI shows the average of the

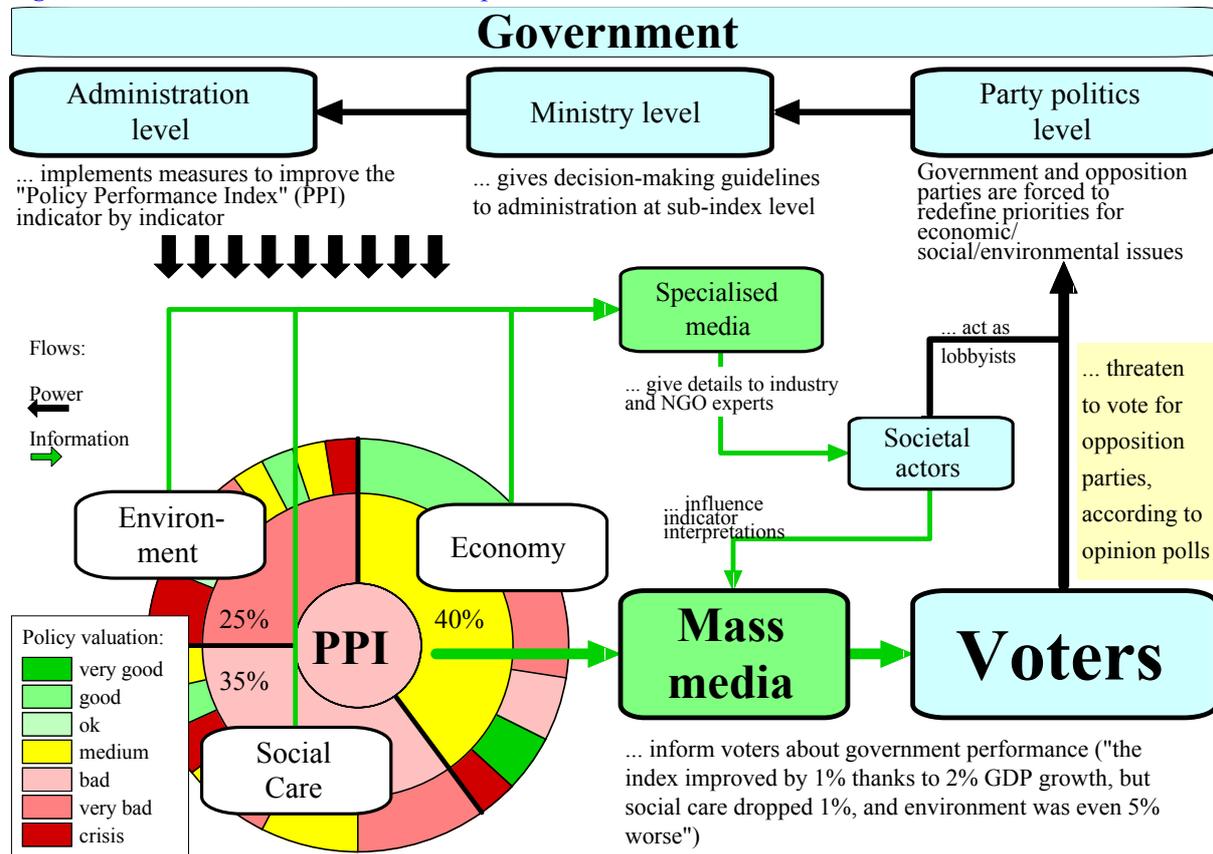


<sup>5</sup> The choice of indicators is illustrative. An interesting question is where to put the unemployment rate: under “Economy” or “Social Issues”? The author believes that unemployment is more a social than a financial problem.

underlying valuations “good+bad+very bad”).

Let us further assume that the Policy Performance Index (PPI) was well-established, and would have substituted GDP and unemployment rate in their role as “headline indicators” for the media. Thus, the citizens would always be able to judge at one glance the government’s performance on a broad range of issues, as displayed above. How would this “perfect” information system influence decision-making? The following figure illustrates (in a very simplified way, of course) the main features of this mechanism:

Figure 4: Indicators, media, voters and politics



Although at first sight the mechanism looks a bit complicated, it provides the government with two simple rules for their decision-making:

- *you must eliminate the red spots*: voters don’t trust governments that are unable to solve a crisis or to deal with a very bad situation;
- *indicators with high weights have a high political priority*: although both the first and the last environmental indicator signal a “crisis”, the government will focus on improving the first one, because it has a higher influence on the colour of the “Environment” segment; and anyway, economic indicators (40%) count more than environmental ones (25%).

Now, this was obviously a look into the future; and some readers, in particular the political scientists among the audience, may rightly say that it is **scary** to imagine that our societies would be driven by a stupid Policy Performance Index, instead of a proper, democratic debate on the priorities of policy-making.

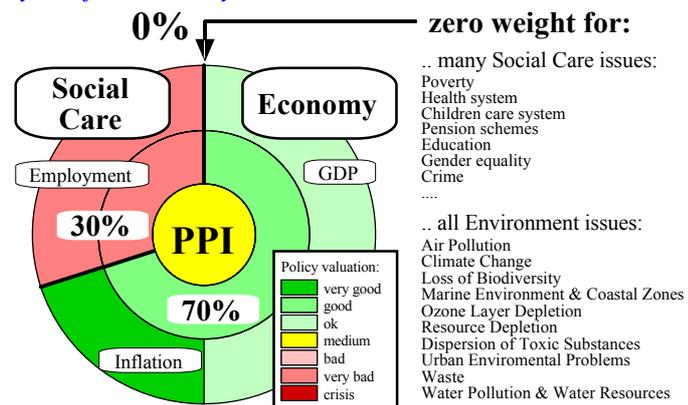
### Where is science in that picture??

(Yes, science is obviously missing in the graph! Not by accident, of course; but we will clarify this important question in the chapter on the role of damage estimates etc. - now we will continue with examining the role of indicators for policy-making, and ignore science for the time being.)

Unfortunately, the mechanisms illustrated in the “*Indicators, Media, Voters and Politics*” figure are in place already today, obviously with a less ambitious information system. Instead of a Policy Performance Index consisting of ca. 20-30 sub-indices, we have **three** indicators (GDP, unemployment rate, inflation rate) that determine - through the voters’ power - the government’s decision-making. Although they are not explicitly aggregated to one PPI, the voter can guess their share in a hypothetical Policy Performance Index by the amount of space dedicated to them in the media <sup>6</sup>; as a regular newspaper reader, I would estimate that the *weight* journalists apply are roughly 50% GDP, 30% unemployment and 20% inflation. The *valuation* currently sees only unemployment as a “red spot”, since in most EU countries GDP grows steadily, and inflation rates are low. These “guesstimates” would result in the following “Policy Performance Index”:

Figure 5: “Reduced complexity” and today’s information system

The “*Indicators, Media, Voters and Politics*” model has always influenced decisions; but today many important societal problems are not covered by established indicators. Although the media will discuss these problems in qualitative terms, the voter has no chance to verify, on the basis of objective indicators, whether the government has performed well on these issues. Accordingly, governments will concentrate their efforts on the *measured* problems.



Instead of being scared by the prospects of a future Policy Performance Index determining the fate of our societies, we should rather look at the disastrous effects of our *current* indicator system, consisting only of a *GDP* that is blind for environmental destruction, an almost irrelevant *inflation* rate, and the *unemployment* rate (which is, although essential and helpful for specific decision-making on the labour market, absolutely insufficient to cover the broad range of social issues that must be addressed by politics).

Given the power of these indicators in the media society, they do not only “reduce the complexity” of the *information* given to the voter; the non-coverage of social and environmental issues by only three economic “headline indicators” **reduces the complexity of real politics**. Governments do not take the decisions that *really* informed voters would expect, but instead concentrate their efforts, willingly or not, on the small fragment of reality that is represented in the limited set of indicators that the media use to report on politics.

<sup>6</sup> Results of an Internet search (Altavista) for the most widely used indicators: GDP >200000, unemployment rate 54000, inflation rate 28000 hits. The most successful competitor to GDP, UNDP’s “Human Development Index” (HDI), had only 7800 hits.

### 3. Better than GDP: a Policy Performance Index (PPI)

#### 3.1 Are we allowed to aggregate “apples and oranges”?

“It is when the hidden decisions are made explicit that the arguments begin. The problem for the years ahead is to work out an acceptable theory of weighting.”

Garrett Hardin, *The Tragedy of the Commons*, Science, Vol. 162, December 1968

Figure 4: *Indicators, media, voters and politics* above shows already how a good solution should look like: an index that covers all important policy issues and communicates policy success or failure to the voter.

What looks so simple at first sight - *aggregate all relevant indicators into one index* - touches one of the oldest disputes of indicator theory: should we “aggregate apples and oranges”? Many experts still categorically reject to do so, saying it is scientifically unsound to compare, for example, car production, hazardous waste and gender equality on the basis of a common unit.

GDP aggregates “apples and oranges” on the basis of their production costs. The resulting Gross Domestic Product is “scientifically *sound*” in the view of its producers, the National Accounting units of Statistical services worldwide. However, when Simon Kuznets and others developed and implemented GDP in the 1940s, their goal was to estimate the capacity of the U.S. economy to sustain an involvement in World War II. Since then, GDP has developed a life of its own. Kuznets himself, and many national accountants, have tried for ages to stop journalists from misusing GDP as a *welfare* indicator, without any success - the *usage* of GDP is “scientifically *unsound*”.

Given its enormous media power, and the devastating distortion of decision-making, the obvious solution would be to **abolish GDP**. However, that would put politicians into the embarrassing situation that they would have to explain their economic policies to the voters in the same way they explain their environmental and social policies today, that is: *without the help of indicators*. It is unlikely that they will agree to abolish GDP; therefore we should try to “heal” the negative consequences of its misuse by reducing its role, and by embedding it into a broader measure of policy performance.

Without repeating all the arguments for and against aggregation <sup>7</sup>, we will examine in the following section a) some essential features of the desired final **product** and b) some of the pitfalls in the **process** towards the ideal “Policy Performance Index”.

#### 3.2 Policy Performance Indices encourage politicians to make good decisions

In a democratic society, the citizen has the right to be informed about politics, so that..

- she or he can build up an own opinion on the importance of a given issue, and on the right way to deal with it (i.e. “*Should income taxes be lowered, if that implies less available money for supporting families with young children?*”);
- she or he can judge if the government acts in the preferred way (i.e. lowers or raises income taxes), so that the voter can decide to re-elect or not this government.

Many nations have laws that guarantee the citizen’s right to be informed; to my knowledge, these laws or directives <sup>8</sup> do not deal with the subtle difference between “information” and

---

<sup>7</sup> see also [http://esl.jrc.it/envind/theory/handb\\_05.htm#Heading30](http://esl.jrc.it/envind/theory/handb_05.htm#Heading30)

<sup>8</sup> Council Directive 90/313/EEC of 7 June 1990 on the freedom of access to information on the environment, Official Journal L 158 , 23/06/1990 p. 56-58, [http://europa.eu.int/eur-lex/en/lif/dat/1990/en\\_390L0313.html](http://europa.eu.int/eur-lex/en/lif/dat/1990/en_390L0313.html)

“communication”. The right to be informed should imply the right to be informed in an understandable way, so that the information is effectively *communicated* to the citizen.

Indicator systems are a means of communication. Beyond pure information, they make complex problems digestible by structuring them, by highlighting what is essential and omitting what is not absolutely necessary to understand a given issue.

A democratic information system, whether consisting of indicators, databases, newspaper columns, TV broadcasts or any other form of reporting and communication, should help the citizen to evaluate the performance of the elected government.

More specifically, a “Policy Performance Index” should *represent*, not *determine*, the perception of importance of a given policy issue. As said earlier, we might be scared by the power of a PPI to drive policy decisions; however, if that power lets politicians take decisions that are in line with their citizens’ expectations, it would be beneficial. A properly designed PPI would do precisely that: it would encourage politicians to dedicate more attention to issues that really matter to people, instead of blindly fighting for more growth.

As illustrated in Figure 4: *Indicators, media, voters and politics*, two main features of a PPI drive policy decisions: the **share** of the respective issue in the index, and the policy **valuation**. If we take the example of GDP (an index measuring production on the basis of a monetary unit), then the car industry would have a higher **share** than the bicycle industry in this index; and for both industries an annual production increase by 5% would be **valued** as a “good” result, while a decrease of 10% within one year would probably be called a “crisis”.

### 3.3 Defining the *share* of the PPI components

While we have clear ideas how to define the respective share of car and bicycle industries in GDP (through their value added measured in Euros), there are no market prices for issues like poverty, gender equality, education, CO<sub>2</sub> emissions or destruction of ecosystems.

There is no easily accessible common unit for these issues; and yet, politicians, when looking at the PPI example above, would probably declare “my friend, you have exaggerated the share of *Environment* a little bit, but I could live with the 35% you attached to *Social Care*”.

We all have a feeling for *importance*, for the weight that such issues have in policy-making. For example, we intuitively know that in Europe *unemployment* is more important than *drugs*, while in the United States it is the other way round. Quantifying such intuition is not too difficult; for example, one could ask the following simple question to a representative sample of citizens:

**Question:** For the purpose of judging the performance of the government, we want to construct a “**Policy Performance Index**”, containing economic, social and environmental indicators. The weight of the indicators should represent the importance of each area for policy-making. If you had 100 points to distribute on the three issues, how many would **you** give to each of them?

(total: 100%)

<b>Economy</b> (e.g. GDP, inflation, investments, ...)	: ___
<b>Social Care</b> (e.g. unemployment, pensions, health system, ...)	: ___
<b>Environment</b> (e.g. climate change, air pollution, waste, noise...)	: ___

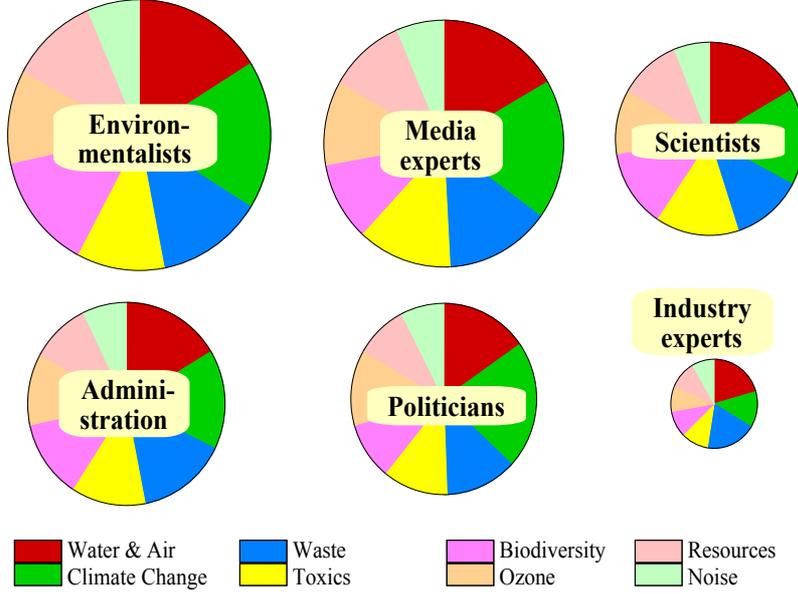
This very straightforward method to determine the weights of an index will work fine as long as the respondent has an opinion on the weight of the issues in real life. An average citizen with a basic knowledge of mathematics who occasionally reads newspapers or watches the

news in TV will be perfectly able to allocate 100 points on economy, social care and environment.

However, the same citizen will have more difficulties, for defining aggregation level 2 of the PPI, to allocate 100 points to **Social Care** issues like *poverty, health system, children care, pension schemes, education, gender equality, drugs and crime* etc. Although it could and should be tried to ask citizens for their opinion on the importance of Social Care issues, one might get more consistent results if the respective question would be asked to a panel of persons working in this policy area; for example, senior experts of the health insurance and pension systems, trade unions, the churches, journalists, doctors, street workers, labour market specialists, and so on (and it will be interesting to compare how the experts perceptions differ from those of ordinary citizens, and why...).

Even more difficult would be the allocation of the 100-point budget on the various components that constitute the policy area “**Environment**”. Again, it could be tried to ask citizens how many points they would give to “Climate Change”, and how many to “Ozone Layer Depletion”. Given that most people do not even understand the difference between the two issues, one should not expect meaningful results. It makes sense to “delegate” the definition of the weights of the environmental sub-index of PPI to a panel of experts who are perfectly familiar with environmental issues. This method was actually tested (using a simple “budget allocation” question) in a 1991 survey<sup>9</sup> among a panel of 660 German senior environment experts, comprising NGO people, journalists, university professors, administrators, politicians (including the members of an environmental Bundestag committee), and industry experts. The results, i.e. the *weight* attached to each of the eight items used, are presented below as pie charts:

Figure 6: Defining the shares of the PPI’s environmental sub-index

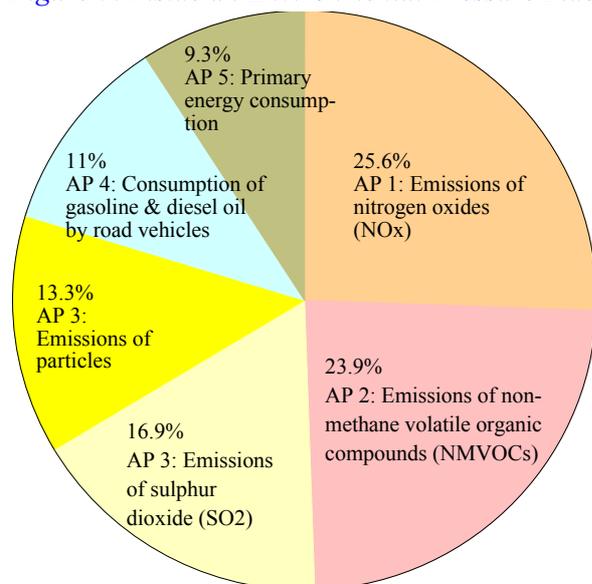


There is a remarkable consensus on the weight of issues even between groups that are “ideologically” far apart, like environmentalists and industry experts. For example, *Climate Change* was consistently given about 50% more weight than the depletion of the *Ozone Layer*.

(Note that in this figure the colours do not represent a valuation - they just serve to distinguish the eight “policy fields” used in this survey. The diameters of the circles show how the expert groups judge the overall judgement of “the” environment.)

<sup>9</sup> conducted by the University of Mannheim (Forschungsstelle für Gesellschaftliche Entwicklungen, FGE)

Figure 7: Inside an Environmental Pressure Index: an Air Pollution Sub-Index



During the definition phase of the Pressure Indices project, a similar question was put forward to the Scientific Advisory Groups (SAG) in autumn 1996:

*“Imagine that you would have to describe the overall pressure in this policy field using a maximum of five **absolutely essential** indicators from the list presented here; which five would you choose?”*

The pie chart represents the average relative weight the sixty European Air Pollution specialists attached to each of the six indicators. Apart from the slightly embarrassing presence of two typical Driving Force indicators (AP 5 & 6) in this pressure index picture, it shows no real surprises - the experts knew what they were talking about.<sup>10</sup>

### 3.4 Defining the **valuation** of the PPI components

#### 3.4.1 Valuation by policy targets?

The objective of a Policy Performance Index (PPI) is to inform the citizen whether the government has done a good or a bad job. Sometimes **policy targets** are being proposed as “anchors” for defining policy success or failure; for example, *if a government promised at the Kyoto summit to stabilise CO<sub>2</sub> emissions at the 1990 levels (or: to raise GDP by 3% per year; to push unemployment below 8%; to increase life expectancy to 99 years; ...), and if the government manages to reach this target, then this should be considered a policy success.*

At first sight, this sounds like a plausible and objective valuation method. However, the “targets approach” suffers from two problems:

- *Which* target should be taken? The Kyoto target certainly has official legitimation, but only a few years ago the Intergovernmental Panel on Climate Change (IPCC) asked, equally legitimated, for a reduction of CO<sub>2</sub> emissions by 75% (which would rightly put the climate policies of *all* UN Member States into the ugly category “complete failure”). The European Environment Agency (EEA) has collected approx. 5,000 targets related to environmental policy - who will define which of them are the “right” and “valid” targets?
- Assuming that we would declare only government targets as “valid” (not such a bad idea because at least EU governments are democratically elected): would an intelligent prime minister ever formulate a target that can only be reached with great sacrifices to the voters? Or would she/he rather declare targets that will be reached anyway with a business-as-usual policy, making thus certain that the PPI segment for the respective indicator (e.g. CO<sub>2</sub> emissions) appears in a dark green (= “very good”) shortly before the elections??<sup>11</sup>

<sup>10</sup> The complete results for all ten policy fields can be found at [http://esl.jrc.it/envind/q2/all\\_q2.htm](http://esl.jrc.it/envind/q2/all_q2.htm)

<sup>11</sup> A closer look at the Kyoto targets will convince the reader that governments chose the second solution.

### 3.4.2 Valuation by benchmarking

In the 1960ies, many European countries had unemployment rates around 1%; inflation was low, and GDP growth was in general higher than nowadays.

In the 1990ies, unemployment reached historical peaks of well over 10% for some countries; inflation was high, and GDP growth was judged “insufficient” by political parties, media and even governments.

If voters had used an “absolute” yardstick for “unemployment performance”, we would have seen victories of the opposition in all elections, given that unemployment rates were ten times higher than in the 1960ies, and given that the newspapers and TV news were dominated by self-appointed economy experts unanimously declaring that GDP growth was too slow.

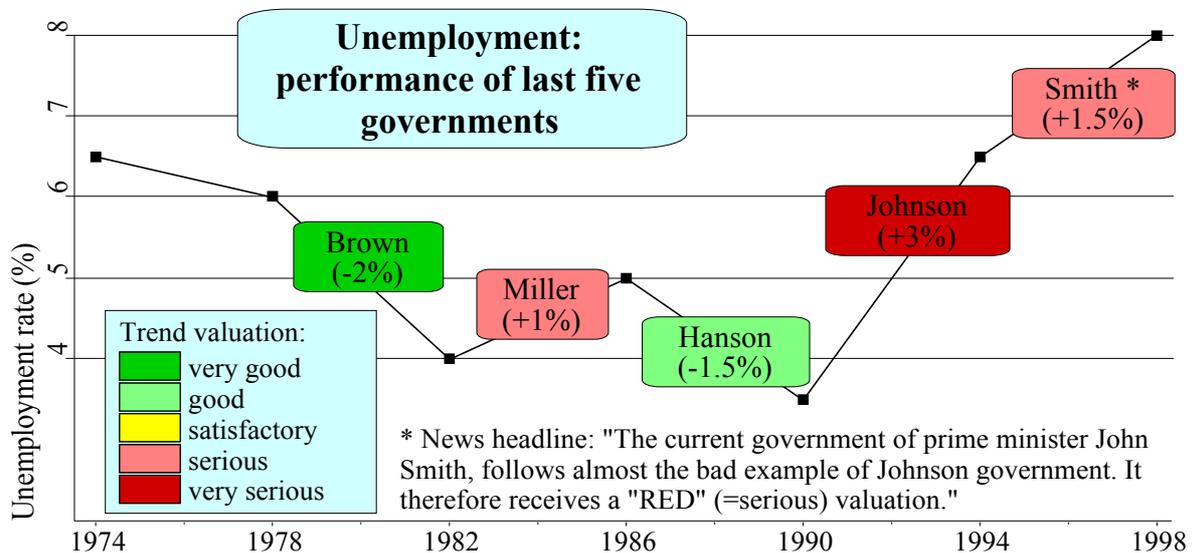
Citizens have a feeling what they can reasonably expect from their governments. The loud propaganda from both sides does not really impress them - they want an objective and *differentiated* picture showing where the weak and strong points of the government are. What they want to know is how the current government performs *relative* to what it *could* achieve under the given constraints; and their yardsticks will usually be:

- how previous governments or opposition parties have dealt with important issues; or
- how the governments of neighbouring countries cope in comparison.

#### 3.4.2.1 Benchmarking I: Comparison to the past

The following illustrative example shows how the current government’s competence in solving the unemployment problem could be compared (for example in a newspaper article) to previous ones. The scale would be delimited by the best (“dark green”) and the worst performance (“dark red”) of the last five governments:

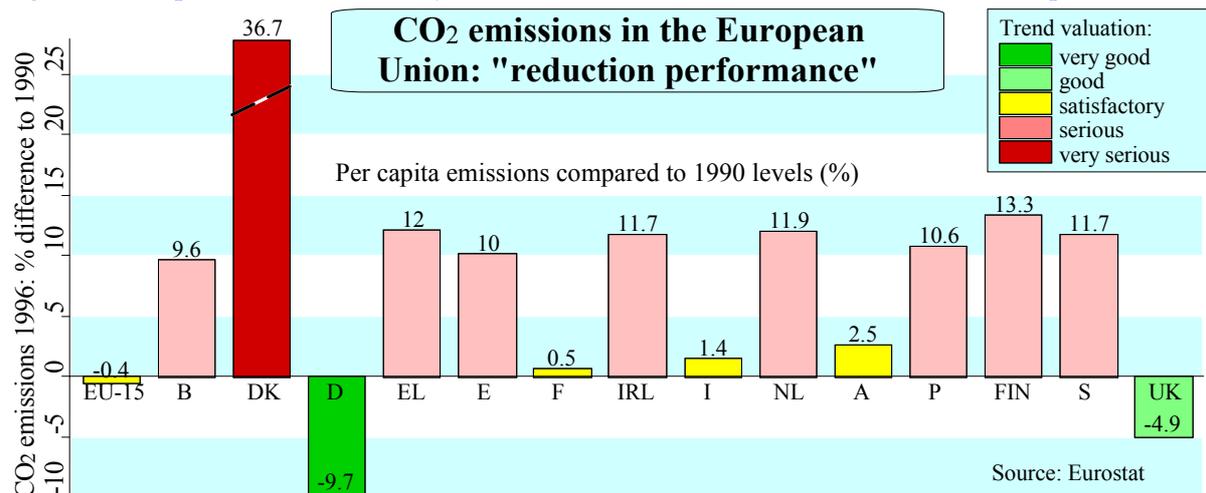
Figure 8: Relative valuation against past policy performance: the unemployment example



#### 3.4.2.2 Benchmarking II: Comparison to the neighbours

As an alternative, the scale for determining the performance of the current government is delimited by the worst and the best of a group of countries, e.g. the fifteen European Union Member States:

Figure 9: Comparison to countries of the same class: the CO<sub>2</sub> emission reductions example

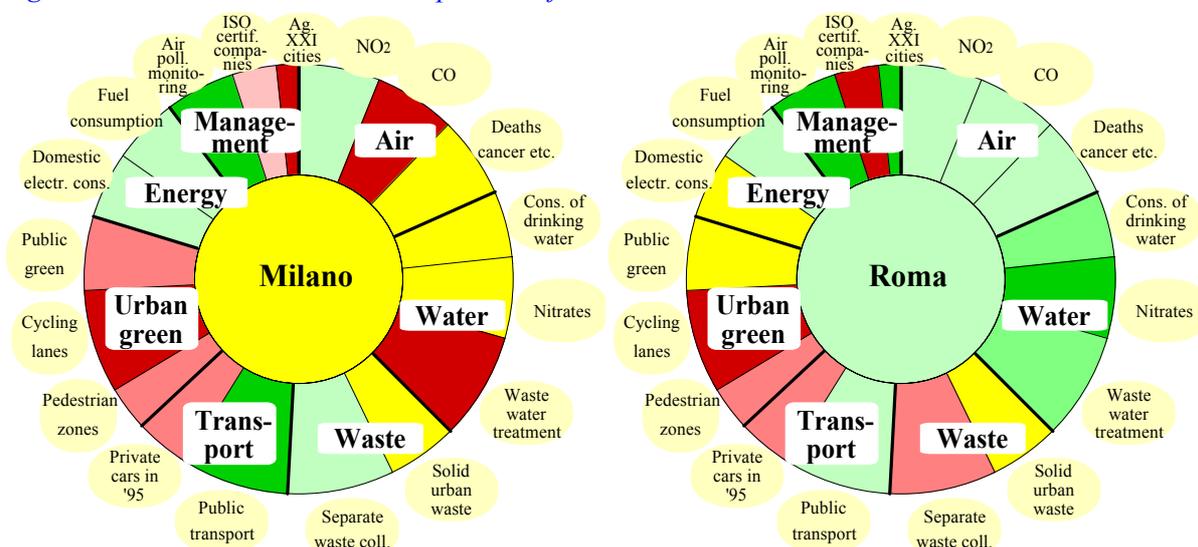


The figure above shows countries' performance with regard to reducing CO<sub>2</sub> emissions in 1996 relative to 1990; the resulting valuations are thus highly policy-relevant for the Kyoto process.

### 3.4.2.3 Benchmarking III: Legambiente's Ecosistema Urbano

At local policy-making level, a remarkable example of a detailed "Policy Performance Index" based on a benchmarking methodology is "Ecosistema Urbano", an environmental index with twenty components developed by Ambiente Italia for Legambiente. The index compares the environmental performance of 103 Italian cities. Below are the results for the "eternal competitors" Rome and Milano (using the presentation scheme introduced above in Chapter 2.3, How indicators influence decisions of citizens):

Figure 10: Ecosistema Urbano: comparison of Italian cities with each other



Ecosistema Urbano has been produced already the fifth time, and is becoming more and more a *management tool* for the cities. The usefulness of this index for decision-making derives from several factors:

- **It reaches the headlines:** Newspapers and journals have published the results, and thus installed a "healthy" competition between the 103 cities. Every mayor knows that next year it will be published again, just like GDP, inflation and unemployment rates.
- **It is transparent:** In contrast to many valuation methods that are accessible only to the expert community, the two figures above are not "black boxes" - everybody can understand the differences between Rome and Milano.

- **It gives policy-makers a chance:** Since the ranking of each single indicator is relative to the other cities, reasonable efforts to solve the problems can lead to significant improvements in the overall ranking; for example, Milano's "red spot" for the indicator "Waste water treatment" will turn into a "green" as soon as the planned treatment plant is operational.

The weights used for constructing the Ecosistema Urbano have been determined. according to Ambiente Italia, in close cooperation with urban administration - a signal that those who are being so merciless judged every year do not perceive the ranking as a threat, but rather appreciate the benefits of public pressure for policy measures aimed to improve the quality of life of their citizens. For a mayor and the urban environment administration, it is easier to defend the expenditure for a waste water treatment plant, a better bus system, or pedestrian zones, if the spending results in a higher rank in the national classification next year.

It is noteworthy that the quality of the underlying statistics (i.e. data coverage and comparability) has considerably improved over the years, demonstrating that steady repetition, transparent methodologies and public attention are the best recipe for getting indicators that are robust enough for politics.

#### 3.4.2.4 Benchmarking IV: UNDP's HDI and the World Economic Forum's ESI

A "benchmarking" system is also the basis for the well-known Human Development Index produced by the United Nations Development Programme (HDI, see HDR99: The Report, at <http://www.undp.org/hdro/HDI.html>), and the recent "Environmental Sustainability Index". Both indices rank a great number of countries against each other. The HDI uses only three indicators, GDP, life expectancy and school enrolment rate. The ESI, produced by the World Economic Forum's "Global Leaders for Tomorrow" (GLT) in cooperation with Yale Center for Environmental Law and Policy and the Center for International Earth Science Information Network (CIESIN) at Columbia University, was released during the Davos summit in January 2000. It ranks 56 countries using **five criteria** and a set of **50** indicators. "The Report also highlights the dire lack of reliable data for many environmental indicators." (Economist, January 29th - February 4th 2000, p. 138).

There are two main **disadvantages** of the "benchmarking" approach to valuation:

- It does not reveal policy failure if all members of a class (e.g. all EU Member States) commit the same errors - for example, not reducing their CO<sub>2</sub> emissions to the levels recommended by the IPCC; one should balance this disadvantage, however, against the political weakness of such far-away targets.
- It requires steady (with regard to comparability over time) and/or internationally compatible indicator sets; such sets exist for OECD and EU Member States (Eurostat and EEA publications), but for Developing Countries progress is slow, and depends strongly on the successful testing and implementation of the UN CSD indicator set.

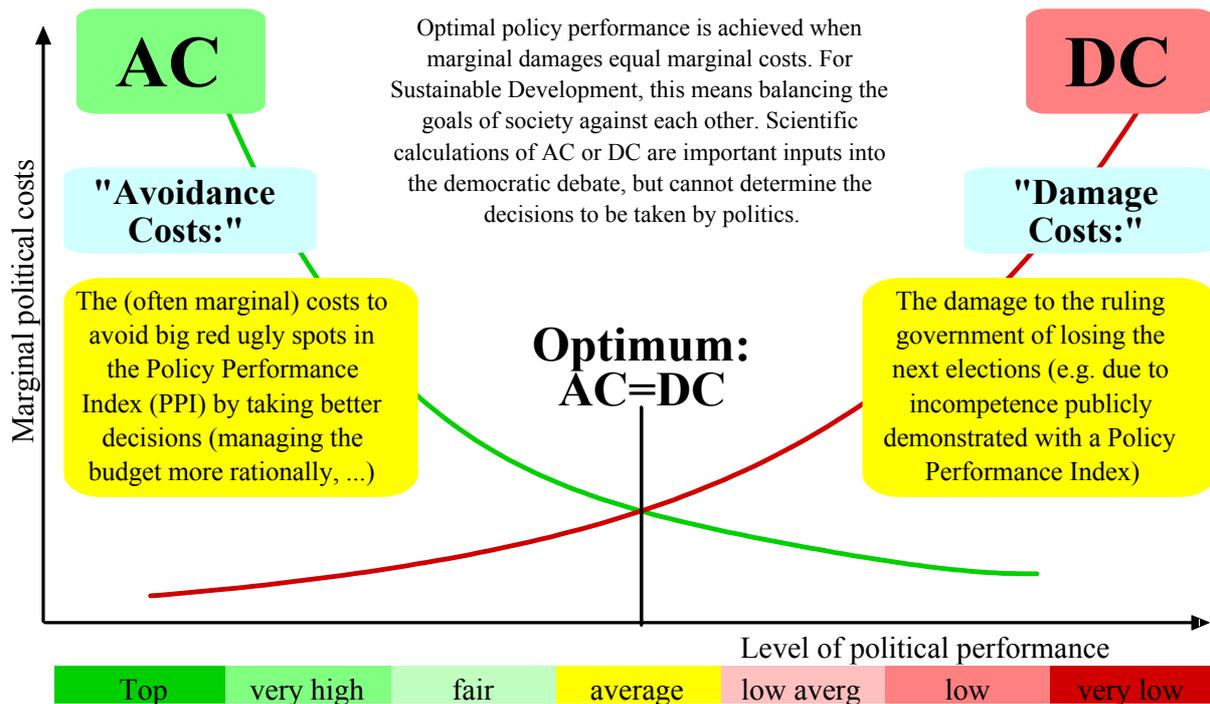
## 4. Costs & Benefits of Sustainable Development Realpolitik

Obviously, the interesting question for this workshop is:

*"Where is the connection to GARP and GreenStamp, and what are the consequences for Environmental Valuation in Europe (EVE)?"*

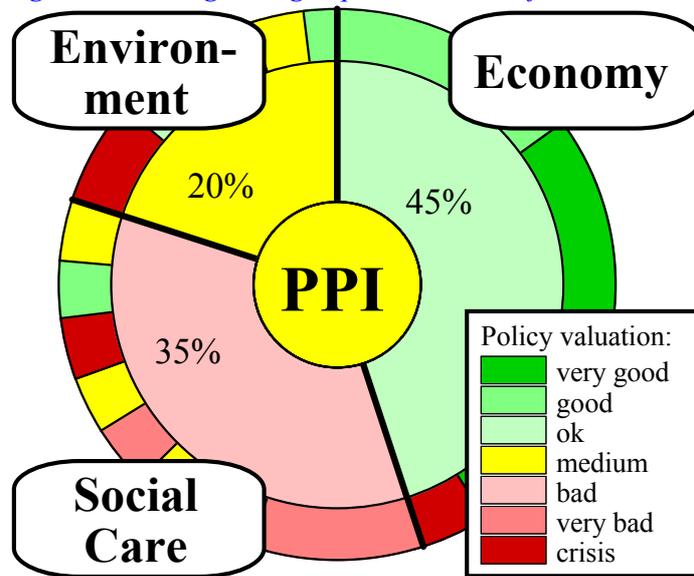
Although GARP and GreenStamp have been launched in an environmental accounting context, their principal questions ("how much are you willing to pay?", "how much will it really cost?") apply also to the more general policy setting described by a Policy Performance Index:

Figure 11: The political dimension of damage and avoidance costs



For serving Sustainable Development policy, both Avoidance Costs and Damage Costs have to be assessed, and the questions to ask are only slightly different than the current ones:

Figure 12: Asking the right questions: WTP for what? Avoiding what?



**The "GARP question":**

How would you allocate the budget of your government on the following three policy areas:

- a) economic growth \_\_\_%
- b) social care \_\_\_%
- c) environment \_\_\_%

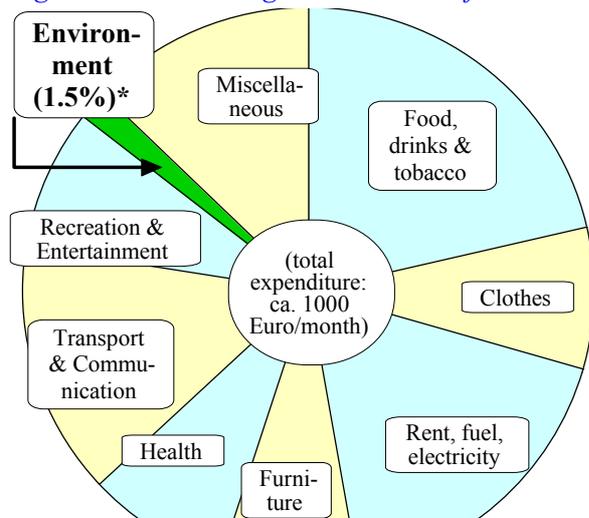
**The "GreenStamp question":**

How much does it cost to improve each of the following three policy areas by one "colour class" (e.g. from "low average" to "average"):

- a) economic growth
- b) social care
- c) environment

## 4.1 How much are you Willing To Sacrifice for a green light?

Figure 13: The average citizen's WTP for the environment



\* The WTP of our societies: only about 1.5% of GDP are being spent for environmental protection. Sources: Eurostat: Basic Statistics of the Community 1994; OECD: PAC Expenditure in OECD Countries 1993

The average citizen's "revealed" Willingness-To-Pay is pretty low: statistics say that our societies spend only about 1-2% of GDP for environmental protection. There are two explanations for this: Either governments act against the voters' preferences, by imposing too "soft" environmental standards, or the WTP is really so low (i.e. people want plenty of food and clothes and entertainment, even if that means a dirty environment...).

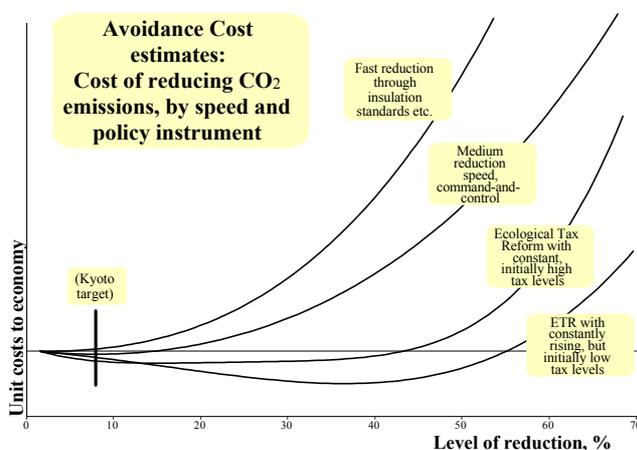
In any case, it would be worthwhile to examine how exactly the WTP for an improvement of a given political issue (i.e. a "red" segment of the Policy Performance Index) is currently being translated into political action, and if the voters' preferences are being respected.

## 4.2 How much does it Cost to Avoid losing the elections?

For Realpolitik, staying in power has an overwhelming importance. Big red segments in a Policy Performance Index signal to the voter "this government is not able to solve a crisis". Accordingly, in order to avoid the catastrophe of being banned to opposition, the government will try to eliminate the red spots, and to turn them into green or at least yellow ones.

The Avoidance Cost approach of GreenStamp could help in formulating the theoretical framework for designing an optimal policy, i.e. one that tries to balance the different goals (material wealth, social care, intact environment) according to society's preferences. In particular, modelling the effects of improving one segment (for example, CO<sub>2</sub> emissions - a sub-segment of the environment cluster) could tell the decision-maker how to design an efficient Climate Change policy that does not immediately meet the resistance of the losers. Obviously, many of the segments of a Policy Performance Index are **linked** with each other; and even for experienced policy-makers it is difficult to keep track of all consequences of a policy option.

Figure 14: Avoidance costs: the CO<sub>2</sub> example



Designing an efficient CO<sub>2</sub> reduction policy is not only a question of comparing damage costs and avoidance costs, and then determine the right standard. Two questions in particular have a great influence on the outcome:

1. Can energy-intensive sectors (i.e. steel industry) be **substituted** with labour-intensive ones (i.e. social services)?
2. How does the **speed** of the implementation influence the costs (i.e. can the "natural" capital substitution be used for a smooth transition, or do companies have to discard costly recent investments)?

Only dynamic models **linking** economy and environment will produce useful answers.

### 4.3 Last but not least: why is there no science in the picture?

Realpolitik needs the support of science, of course; but will politicians ever react to a statement like the following?

*“During a recent workshop, the researchers of the GARP and GreenStamp project declared that there is a huge gap between the welfare loss induced by CO<sub>2</sub> emissions, and the costs to avoid this loss. According to the project leaders, the point where Damage Costs and Avoidance Costs meet corresponds to a CO<sub>2</sub> reduction of 65% over the next five years.”*

It would be nice if the prime minister reacted immediately to such brilliant science, but experience tells us that the one million pages that were written for the preparation of the UNCED conference in 1992 (full of “science”, we assume) had absolutely NO impact on the world’s CO<sub>2</sub> emissions.

Unfortunately, Realpolitik has never paid great attention to scientific evidence. Politicians react when they are sufficiently scared of losing voters; but voters are not scientists - they depend entirely on second-hand information. Being flooded by such information, they are grateful to dispose of a handful of “objective” indicators - see above. However, the weight given to the segments of the PPI will never be determined by “science”, but rather by the slow and difficult process of feeding scientific results into the public debate. A consistent policy framework, displayed as a Policy Performance Index, may at least help to structure the debate, and to give an equal voice to economists, environmental and social sciences.

## 5. Annex: Basic Guide for Constructing a Policy Performance Index

### 5.1 How many issues?

Most proposals for SD indices present three main issues: economy, environment, social issues. It could be argued that the latter does not cover cultural and political aspects, and that a fourth issue (“culture”?) is needed.

### 5.2 Which issue sets?

The three or four issues should be given names that respond to people’s expectations and at the same time are widely applicable within the chosen context (e.g. comparison of OECD countries; see also [www.esl.jrc.it/envind/theory/handb\\_07.htm](http://www.esl.jrc.it/envind/theory/handb_07.htm) for a discussion of the issue definition in an environmental policy context).

### 5.3 How many indicators per issue?

The level of detail needed to ensure that the most important political issues are covered by the index determines its credibility. Low detail (the HDI, for example, has only one indicator per issue) makes the index simpler and easy to introduce, but limits its use. Too high detail will make it impossible.

### 5.4 Which indicators?

Let us assume we had decided that we need 10 indicators for the issue “social care”. Does that include *unemployment*, or should *unemployment* be placed under “economics”? Obviously, such questions are not trivial and require some kind of selection process. Choosing the sixty indicators of the European Pressure Indices project involved 2,300 EU experts and cost a lot of money. Given that many of these indicators had to be produced for the first time, and that 15 EU states had to be involved, the expert survey was still a transparent and efficient way to speed up the discussion on the right indicators. However, especially if many candidate indicators are readily available, other methods to define the set could be applied, like counting the number of times journalists have reported on issues that could be quantified by indicators. The main difficulty in the crucial step of indicator definition is to accept that it should not be

determined by one's own bright ideas, but rather *represent* the mainstream of societal debates as objectively as possible. <sup>12</sup>

### 5.5 Which weights?

Starting with the assumption that all ten selected indicators should be given a 10% share in the “social care” sub-index would shorten the debate, but sooner or later somebody will notice that, e.g. in a Developing Countries context, “income distribution” is more important than “access to safe drinking water” - and a more sophisticated weighting procedure will be needed, using for example the “budget allocation process” described above. One main pitfall of weighting is the risk to assign a weight of zero to an important political issue - by not including it in the list of indicators in the *previous* step.

### 5.6 How to value a trend as “policy success” or “policy failure”?

In principle, one could use the comparison to past governments or to neighbouring states, as described above. Often the comparison to the past is difficult, due to the lack of time series. In practice, the comparison between countries, regions or cities is easier, as demonstrated by the examples Human Development Index (HDI), Environmental Sustainability Index (ESI) and Ecosistema Urbano.

---

<sup>12</sup> Should “number of olympic medals won by my country” be included in the cluster “cultural performance”? Personally, I would strongly disagree - but the Policy Performance Index should represent the interests and concerns of **all** citizens, and therefore my personal opinion as an indicator expert should not count at all.