

June 2, 2009

COMMISSION ON THE MEASUREMENT OF ECONOMIC PERFORMANCE AND SOCIAL PROGRESS

DRAFT SUMMARY

(PROVISIONAL AND INCOMPLETE)

1. *The Commission on the Measurement of Economic Performance and Social Progress (CMEPSP, also called “the commission” in the following text) has been launched in February 2008 by the President of the French Republic. It started its work at the end of April 2008. After one year, the commission has nearly completed its tas, but its members are still working on issues raised during former plenary meetings or subgroup meetings. The present document is hence a provisional summary of the commission’s work. Its aim is to collect comments and suggestions from the civil society at a time when it is still possible to consider them.*
2. *If readers of the present draft summary want to communicate comments and/or suggestions to the commission, they have to click on the place indicated at the end of the homepage of the commission. They will access an e-mail form where to write and send their comments and/or suggestions to the secretariat of the commission. The deadline is July 5, 2009.*

TABLE OF CONTENTS

INTRODUCTION	3
CHAPTER 1 – CLASSICAL GDP ISSUES	6
1. GDP, prices and living standards	6
2. Imputations – comprehensiveness versus comprehensibility	9
3. What can be done within the existing measurement framework?	11
3.1. Emphasize national accounts aggregates other than GDP	11
3.2. Improving the measurement of government-provided services	14
3.3. Revisit the concept of ‘defensive’ expenditures	17
3.4. Income and wealth come together	19
4. Bringing out the household perspective	21
4.1. Adjusting household income measures for government services in kind	21
4.2. Medians and means - distribution of market income	24
4.3. Broader measures of households’ economic activity	29
4.4. Distribution of full income	35
CHAPTER 2 – QUALITY OF LIFE	36
1. Introduction	36
2. Conceptual approaches to measuring quality of life	37
3. Subjective measures of quality of life	39
4. Objective features shaping quality of life	42
4.1. Health	42
4.2. Education	44
4.3. Personal activities	46
4.4. Political voice and governance	49
4.5. Social connections	51
4.6. Environmental conditions	52
4.7. Personal insecurity	54
4.8. Economic insecurity	55
5. Cross-cutting issues	57
5.1. Assessing links across quality of life dimensions	57
5.2. Inequalities in quality of life	58
5.3. Aggregating across quality of life dimensions	60
CHAPTER 3 – SUSTAINABLE DEVELOPMENT AND ENVIRONMENT	62
1. Introduction	62
2. Taking stock	64
2.1. Dashboards or sets of indicators	65
2.2. Composite indexes	67
2.3. Adjusted GDPs	69
2.4. Indexes focusing on overconsumption or underinvestment	74
2.4.1. Adjusted net savings (ANS)	74
2.4.2. Footprints	78
3. Summarizing sustainability with a single number: why is it so difficult ?	81
3.1. What do we want to measure ?	81
3.2. What would be the requirements of an appropriate index ?	84
3.3. Technological uncertainties argue in favor of a more hybrid approach	86
3.4. Uncertainty is also normative	88
3.5. An additional source of complexity: the transnational dimension	90

INTRODUCTION

3. The Issues Paper produced by the commission in July 2008 has played an important role in the organization of the work of the commission. It underlined the existence of a huge distance between standard measures of important socio economic variables like growth, inflation, inequalities, etc. and widespread perceptions by the population; the gap is so large and so universal that it cannot be explained by reference to money illusion and/or to psychological characteristics of human nature. Our statistical apparatus, which may have served us well in a not too distant past, is in need of serious revisions.
4. The creation of the commission is reflecting this concern. As its name suggests, its focus is on the measurement of economic performance and social progress. Currently, the most widely used metric is GDP (gross domestic product.) Its aim is to identify the limits of GDP as an indicator of economic performance and social progress, to consider additional information required for the production of more relevant indicators, to discuss how to present this information in an appropriate way, and to assess the feasibility of alternative measurement tools.
5. Indeed, for a long time there have been increasing concerns about the adequacy of current measures of economic performance, in particular those based on GDP figures. Moreover, there are even broader concerns about the relevance of these figures as measures of societal well-being. The inadequacies of these figures from the perspective of *sustainability*-- economic, environmental, and social sustainability-- have been of particular concern.

6. One outcome of the Commission's work will be suggestions for alternative indicators which may provide a better description of economic performance and social progress. Taking stock of similar work conducted in the past, the Commission will be cautious about the number of indicators proposed. Here, as elsewhere in economics, there are trade-offs: a larger number of indicators may better reflect the diversity of issues and individual situations, but an excessively large number may provide a confused picture of the overall situation. On the other hand, a single figure mixing a large number of socio-economic phenomena provides an inadequate basis for appropriate policy measures.
7. Any statistical indicator has to aggregate variables that are, in some sense, incommensurate. In estimating GDP, we add up apples and oranges; and we aggregate them together using *relative* prices. If an orange sells for twice the price of an apple, then each orange is counted as two apples. The justification of this is that *in competitive markets*, relative prices reflect marginal relative valuations. An orange is valued by *all* consumers as "worth" twice as much as an apple. While even in market transactions, this assumption may be questioned (for instance, when markets are imperfect), when moving *beyond* GDP to areas where there are limited or no market transactions, the relevance of a *monetary* metrics becomes more questionable. The choice of alternative metrics has to be assessed both from a conceptual and practical point of view.
8. To organize its work, the Commission selected three main directions of study which correspond to three of the already identified main causes of divergences between perceptions and measures. (i) *Classical GDP issues*. Limits of GDP as an indicator of socio-economic progress or economic performance can be addressed by investigating possible extensions or modifications of the current conceptual framework; (ii) *Quality of life*. This direction of study is approaching the measurement of social progress from a broader perspectives on *well-being*, taking also into account metrics derived from asking

people about how they themselves feel; (iii) *Sustainable development and environment*.

As noted above, one of the biggest concerns about current measures of economic performance and social progress is related to sustainability and one of the areas where sustainability is most questioned is the environment.

9. The present *Draft Summary* is accordingly structured into three summary chapters based on the work of the three sub-groups created by the Commission on the occasion of its first plenary meeting:

- Classical GDP issues
- Quality of life
- Sustainable development and environment

CHAPTER 1 – CLASSICAL GDP ISSUES

1. GDP, prices and living standards

10. Gross domestic product (GDP) is the most widely used measure of economic activity. There are international standards for its calculation and much thought has gone into its statistical and conceptual bases. But GDP is a measure of mainly *market* production, though it has often been treated as if it were a measure of economic well-being. Doing so can lead to misleading indications about how well-off people are and entail the wrong policy decisions.

11. One of the reasons why money measures of performance and living standards have come to play such an important role in our societies is that monetary valuation of goods and services makes it easy to add up quantities of a very different nature. When we know the prices of apple juice and of DVD players, we can add up the values of apple juice and DVD players and make statements about production and consumption in a single figure. But market prices are more than an accounting device. Economic theory tells us that when markets are functioning, the ratio of one market price to another is reflective of the relative appreciation of the two products by those who purchase them. GDP captures all final goods in the economy, whether they are consumed by households, firms or government. Valuing them with their prices would thus seem to be a good way of capturing, in a single number, how well-off society is at a particular moment. And keeping prices unchanged while observing how quantities of goods and services that enter GDP move over time would seem like a reasonable way of making a statement about how society's living standards evolve in real terms.

12. As it turns out, things are more complicated. First, prices may not exist for some goods and services (if for instance government provides free health insurance or if households are engaged in child care) and there is a question how to value these services. Second, even where there are market prices, they may deviate from society's underlying valuation. In particular, when the acts of consumption or production of individuals affect society as a whole, the price that individuals pay for a product will differ from its value to society at large. Environmental damage caused by production or consumption activities that is not reflected in market prices is a well-known example.
13. There is yet another problem. While it is straight forward to talk about 'prices' and 'quantities' in concept, defining and measuring price and quantity changes in practice is altogether a different matter. As it happens, many products change over time – they disappear entirely or there are new features added to them. *Quality change* can be very rapid in areas like information and communication technologies. And there are products whose quality is complex, multi-dimensional and hard to measure such as medical services, education services, research activities or financial services. There are also problems associated with the collection of data, in an era where an increasing fraction of sales are done over the internet and at sales as well as discount stores. As a consequence, capturing quality change correctly is a tremendous challenge for statisticians and yet it is vital to measuring real income and real consumption, some of the key determinants of people's well-being. Under-estimating quality improvements is equivalent to over-estimating the rate of inflation and therefore equivalent to under-estimating real income. For instance, in the mid-90s, a report reviewing the measurement of inflation in the United States (Boskin Commission Report) estimated that insufficient accounting for quality improvement of goods and services led to an annual over-estimation of inflation by 0.6%.

14. The debate in Europe has tended to go the opposite way: official price statistics have been criticized for *under-estimating* inflation. Partly, this has been because people's perception of inflation differs from the national averages that are presented in the consumer price index. Partly, this has been because it is felt that statisticians over-adjust for quality improvements of products, thereby painting too rosy a picture of the real income situation of citizens.
15. For market prices to be reflective of consumer's appreciation of goods and services it is also necessary that consumers are free to choose and dispose of the necessary information. It takes little imagination to argue that this is not always the case. Complex financial products are an example where consumer ignorance prevents market prices from playing their role as carriers of correct economic signals. Complex and ever-changing bundles of services offered by telecommunication companies are another case in point where it is difficult to ensure transparency and comparability of price signals.
16. All the above considerations imply that in temporal and spatial comparisons, price signals have to be interpreted with care. For a number of purposes, they do not provide a useful vehicle for aggregation of quantities. This does not imply that the use of market prices in constructing measures of economic performance is generally flawed. But it suggests prudence, in particular with regard to the single, and often over-emphasized measure, GDP.
17. This Chapter suggests five ways of dealing with some of the deficiencies of GDP as an indicator of living standards. First, inside the national accounts emphasize other, and normally well-established - indicators than GDP. Second, improve the empirical measurement of key production activities, in particular the provision of health and education services. Third, bring out the household perspective which is most pertinent for

considerations of living standards. Fourth, add information about income and wealth distribution to data on the average evolution of income and wealth. Finally, we can think about widening the scope of what is being measured. In particular, a significant part of economic activity happens outside markets and is often not reflected in established national accounts. However, when there are no markets, there are no market prices and valuing such activities requires estimates ('imputations'). These are meaningful but they come at a cost, and we shall discuss them before turning to the other proposals...

2. Imputations – comprehensiveness versus comprehensibility

18. Imputations exist for two related reasons. First is for reasons of comprehensiveness. There are productive activities and associated income flows (typically non-monetary) that take place outside the market sphere and some of them have been incorporated into GDP. The single most important imputation is a consumption value for the services that home owners derive from living in their own dwellings. There is no market transaction and no payment takes place but the national accounts treat this situation as if home-owners paid a rent to themselves. Most people would agree that two persons with the same money income but one of them living in his/her own house and the other renting are not equally well off - hence the imputation to better compare incomes over time or between countries. This brings us to the second reason for imputations, the *invariance principle*: the value of main accounting aggregates should not depend on the institutional arrangements in a country. For example, if exactly the same medical services are provided in one case by the public sector and in another case by the private sector, overall measures of production should be unaffected by a switch between the two institutional settings. The main advantage of adhering to the invariance principle is better comparability, over time and

between countries. Therefore, for instance, measures of ‘adjusted disposable income’ for households (see below) include an imputation for government services provided directly to citizens.

19. Imputations are more or less sizeable, depending on the country and on the national accounts aggregate considered. The table below indicates that the main imputations account for about one third of adjusted disposable income of households in the two European countries and for just over 20% in the United States, with a rising tendency in all three countries. Thus, the living standards of French and Finnish households would be understated relative to the United States in the absence of imputations.

Table 1: Major imputed and non-imputed components of adjusted disposable household income

	France		USA		Finland	
	1985	2007	1985	2007	1985	2006
Imputed rents	6.9%	10.1%	8.8%	10.1%	9.2%	12.2%
Financial services including FISIM	3.3%	1.5%	2.9%	4.3%	1.9%	2.3%
Social transfers in kind	17.3%	19.0%	7.9%	8.3%	19.5%	22.3%
Total imputations	27.4%	30.6%	19.6%	22.8%	30.6%	36.8%
Other disposable income (not imputed)	72.6%	69.4%	80.4%	77.2%	69.4%	63.2%
Total adjusted disposable income	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Source: OECD annual national accounts.

20. But imputations come at a price. One is data quality: imputed values tend to be less reliable than observed values because they require assumptions about implicit transactions. The other is the effect of imputations on the comprehensibility of national accounts. Not all imputations are perceived as income-equivalent by people and the result may be a discrepancy between the changes in perceived and the changes in measured income. This problem is reinforced when we widen the scope of economic activity to include other services that are not mediated by the market. Our estimates below for household work amount to around 30% of conventionally-measured GDP. And another

80% or so are added when leisure is valued as well. It is undesirable to have assumption-driven data so massively influencing overall aggregates.

21. There is no easy way out of the tension between comprehensiveness and comprehensibility except keeping *both* elements of information available for users and keeping a distinction between core and satellite accounts. A full set of household accounts, for example, may not be well placed in the core of national accounts aggregates. But a satellite account that comes up with a valuation of comprehensive forms of household production can be useful information.

3. What can be done within the existing measurement framework?

3.1. Emphasize national accounts aggregates other than GDP

22. A first step towards mitigating some of the criticism of GDP as a measure of living standards is to emphasize national accounts aggregates other than GDP. For example, one would like to account for depreciation and thus deal with *net rather than gross measures* of economic activity.
23. Gross measures take no account of depreciation of capital goods. If a large amount of output produced has to be set aside to renew machines and other capital goods, society's ability to consume is less than it would have been if only a small amount of set-aside were needed. The reason that economists have relied more heavily on GDP than on net domestic product (NDP) is, in part, that depreciation is hard to estimate. When the structure of production remains the same, GDP and NDP move closely together. But in recent years, the structure of production has changed. Information technology (IT) assets have gained importance as capital goods. Life expectancy of computers and software is

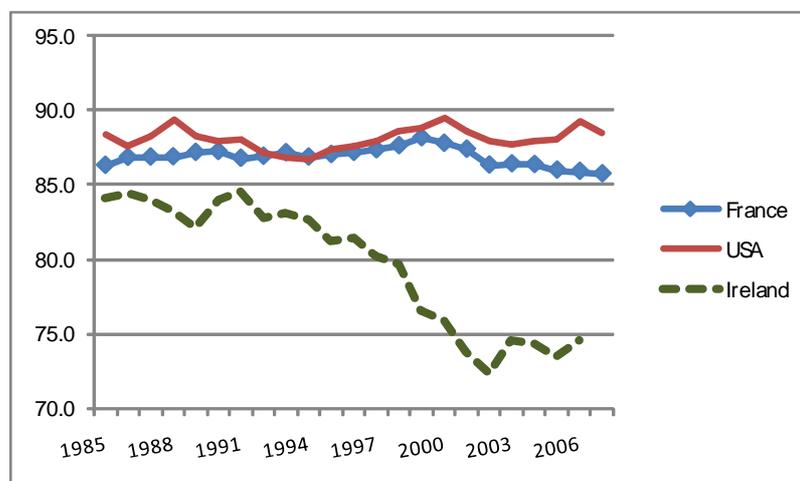
shorter than of steel mills. On those grounds, the discrepancy between GDP and NDP may be increasing, and by implication, volume NDP may be increasing less rapidly than GDP. For example, real GDP in the United States rose by about 3% per year during the period 1985-2007. Depreciation rose by 4.4% over the same period. As a consequence, real net national product grows at a slower rate than GDP.

24. Of greater concern for some countries is that the standard depreciation measures have not taken into account the depletion of scarce natural resources, and the degradation in quality of the natural environment. There have been various attempts to widen the scope of depreciation to reflect environmental degradation (or improvement if such is the case) but without much success. The hurdle is reliable measurement. In market-based accounting, market prices are used for valuation. Thus, a ton of coal extracted is valued by the market price for coal. But the main point of introducing environmental assets is to alert to the social costs of the consequences of a ton of coal extracted, and these are not normally reflected in market prices for coal. Taking into account resource depletion would, for instance, suggest a smaller role in NDP for sectors like coal mining and timber than in GDP, and in some instance, there has been lobbying against the development of more comprehensive accounts that would reflect resource depletion, and especially adverse environmental effects.

25. In a world of globalization, there may be large differences between *income* of the citizens of a country and measures of domestic *production*, and the former is clearly more relevant for measuring the well being of citizens. We shall argue later that the household sector is particularly relevant for our considerations and for households, the income perspective is much more appropriate than measures of production. Some of the income generated by residents is sent abroad, and some residents receive income from abroad. These flows are captured by *net national disposable income*, a standard variable in

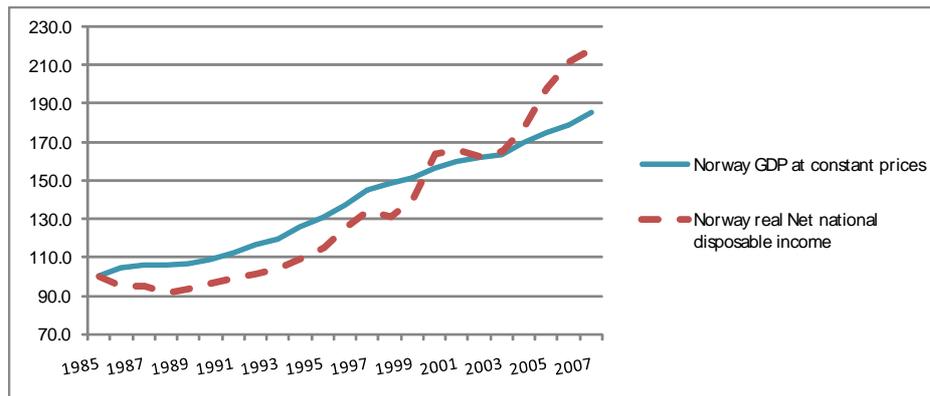
countries' national accounts. The figure below shows how Ireland's income declines relative to its GDP – a reflection of an increasing share of profits that are re-patriated by foreign investors. While the profits are included in GDP, these profits do not enhance the spending power of the citizens of the country. For a poor developing country to be told that its GDP has gone up may be of little relevance. It wants to know if its citizens are better off and national income measures are more relevant to this question than GDP.

Figure 1: Net national disposable income as percentage of gross domestic product



26. Moreover, prices of imports evolve very differently from the prices of a country's exports, and these changes in relative prices have to be taken into account in assessing living standards. The figure below shows the divergence between income and production for Norway, an oil-rich OECD country whose income has risen faster than GDP in times of rising oil prices. In many developing countries, whose export prices fall relative to the prices of imports, the opposite will be true. These effects are captured by measures of *real net national disposable income*, available from most countries' national accounts but much more rarely used in the public debate than GDP.

Figure 2: GDP and disposable income in Norway



Source: OECD Annual National Accounts.

3.2. Improving the measurement of government-provided services

27. Governments play an important part in today's economies. Broadly speaking, they provide two types of services – those of a 'collective' nature such as security and those of an 'individual' nature such as medical services or education. This does not imply that government is necessarily the only provider of these services and indeed, the mix between private and public provision of individual services varies significantly across countries. And while one can argue about the contributions of collective services to citizens' living standards, individual services – education, medical services or public sports facilities – are almost certainly positively valued by citizens. These services tend to be important in size and at the same time badly measured. Traditionally, for government-provided non-market services, measures have been based on the *inputs* used to produce these services rather than on the actual *outputs* produced. An immediate consequence of this procedure is that multifactor productivity change for government-provided services is ignored because outputs are taken to move at the same rhythm as inputs. It follows that if there is faster productivity growth in the public sector than in the private, our measure under-estimates growth, and conversely if there is slower growth.

28. Work has started in many countries to develop output measures that are independent from inputs but the task is tremendous. Take the following example: the United States spend more per capita on health care, and yet in terms of standard health indicators, health outcomes are worse than in many European countries. Does this mean that Americans get less health care? Or does it mean their health care is more expensive and/or delivered less efficiently? Or does it mean that health outcomes also depend on factors specific to the American society other than health expenditures? We need to be able to break the change in health expenditures into a price and an output effect. But what exactly are the volumes of output that one is looking for? It is tempting to measure them by the state of the health of the population. The problem is that the link between expenditures on health care and health status is tenuous at best: expenditures relate to the resources that go into the institutions providing health and education services whereas the health status of the population is driven by many factors, all of which together form the health or education *system*. For example, people's lifestyle will affect health outcomes or the time parents spend with their children will affect exam scores. Attributing changes in the health or education status only to hospitals and schools and the money spent on them neglects all these factors and can be misleading.

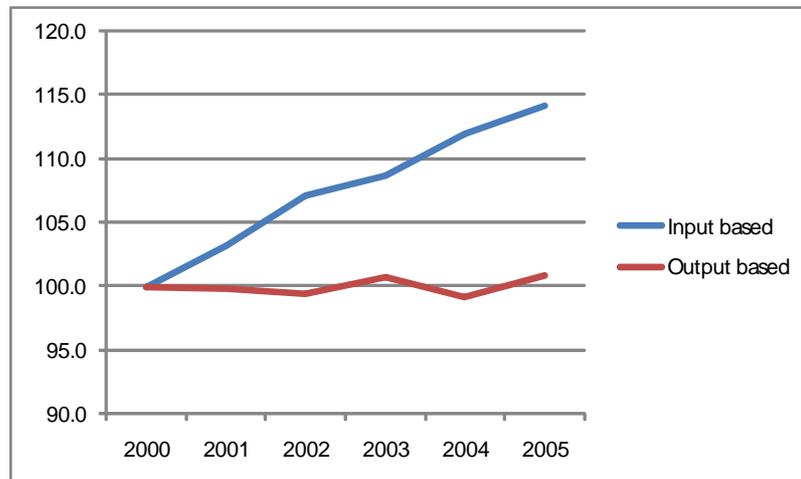
29. The quest is for more accurate measures of the volume growth of public services. A number of European countries as well as Australia and New Zealand have developed output-based measures for key government-provided services. A major challenge in these efforts is again capturing quality change. Without a good measure of quality (or equivalently, a good estimate of the increase in productivity) it is impossible to ascertain whether the conventional input measures under- or overestimate growth. If undifferentiated quantity measures such as a simple number of students or of patients are used, changes in the composition of output and in its quality may be missed. But one has

to start somewhere and the issue cannot be ignored because the numbers involved are important. For example, with output-based measures, the U.K. economy grew at the rate of 2.75% per year between 1995 and 2003, whereas if the previous convention had continued to be used, the growth rate would have been 3% (Atkinson 2005). Similar effects could be observed in the case of France. A Danish study on measurement of health output points the other way: output-based prices of hospital services grew less rapidly than input-based prices (Figure 3), indicating that real growth of medical services has been understated.

30. An important criterion for the reliability of output-based measures is that they are based on observations that are detailed enough to avoid obfuscation of the results because true volume changes are mixed up with compositional effects. We can ask: how many students are educated and simply count their numbers. If spending per student increases, one might conclude that the unit cost of education services has increased. This may, however, be misleading if costs have gone up because students are taught in smaller classes. Or if there is a larger share of students that take up more costly engineering studies. The measurement mistake is that the simple number of students is too undifferentiated an output measure to be meaningful. A more detailed structure is needed for measurement. This helps treating for instance one hour taught to a graduate engineering student as a different product from one hour taught to a first year student in arts. In so doing, we manage to account for some quality and compositional change. A similar reasoning applies for health care where the treatments of different diseases have to be considered as different medical services. As it turns out, health care policies in some countries have actually helped making available the administrative data needed to obtain this detailed information. We conclude that despite it being a daunting task, *better measurement of government-provided individual services is central* to our assessment of

living standards. *Exploitation of new administrative sources to deepen the level of information detail* is one way of progressing in this direction.

Figure 3: Unit cost indices for general hospitals in Denmark



Source: Deveci, Heurlén and Sørensen (2008).

3.3. Revisit the concept of ‘defensive’ expenditures

31. Expenditures required to maintain consumption levels or the functioning of society could be viewed as a sort of intermediate inputs – there is no direct benefit and in this sense they do not give rise to a final good or service. Many such ‘defensive expenditures’ are incurred by government, others are incurred by the private sector. By way of example, expenditure on prisons could be considered a government-incurred defensive expenditure, and costs of commuting to work serves as an example for typically privately-incurred defensive expenditures. A number of authors suggested treating these expenditures as intermediate rather than final products. Consequently, they would not be part of GDP.
32. Defensive expenditures not only concern government activity. Nordhaus and Tobin’s (1973), for example, identify as ‘defensive’ those activities that "are evidently not directly sources of utility themselves but are regrettably necessary inputs to activities that may

yield utility". In particular, they adjust income downwards for expenditures that arise as a consequence of urbanization and a complex modern life.

33. At the same time, difficulties abound when it comes to identifying which expenditures are 'defensive' and which are not. What are possible ways forward? Options include:

- *First, focus on household consumption rather than total final consumption.* For many purposes, this can be a meaningful variable. And all of governments' collective consumption expenditures (which would include things like prisons, military expenditure or the clean-up of oil spills), are automatically excluded from households' final consumption.

- *Second, widen the asset boundary.* In many cases, there are elements of *investment*, and *capital* goods, and in those cases, we can treat defensive expenditures much like maintenance expenditures in the case of conventional production. Expenditure on security could be taken as an investment in social capital, akin to health expenditures that can be seen as investment in human capital. If there is an asset that captures environmental quality, the consequences of economic activity that is detrimental to this asset could be captured in an extended measure of depreciation or depletion so that the *net* measure of income or production is reduced accordingly. And net measures, it was argued earlier, should be our benchmark for living standards rather than a gross measure.

- *Third, widen the household production boundary.* Some 'defensive' expenditures cannot reasonably be treated as an investment. Take the case of commuting to work. Households produce transportation services – they use their time (labour input) and money (commuter ticket) for this purpose. These transportation services are provided to employers for whom they constitute a free intermediate input. With

the exception of the consumer's purchase of the ticket for a commuter train, none of the above flows enter measures of production and income. This could be remedied by allowing for household production of transportation services and simultaneously imputing a flow of intermediate inputs from household producers to firms. The additional imputed purchase would reduce firm's measured profits by the amount of transportation services produced by households with offsetting effects on overall income. However, commuter train tickets would now figure as intermediate inputs into household production rather than as final household consumption and the value of GDP under this new treatment would be lower by exactly the defensive expenditure 'commuter ticket' than the value of GDP under present accounting practice.

34. The biggest obstacle to these approaches lies in their implementation. How exactly should the scope of defensive expenditures be determined? How should the new assets and in-kind flows be valued? And of course, widening the scope of asset and production measures brings with it more imputations.

3.4. Income and wealth come together

35. Income flows are an important gauge for the standard of living but in the end, it is wealth that determines how well off people are. The best indicator of the financial status of a firm is its balance sheet, and the same holds for the economy as a whole. To construct the balance sheet of an economy, we have to have comprehensive accounts of its assets (physical capital - and perhaps even human, natural and social capital) and its liabilities (what is owed to other countries.) To know what is happening to the economy, we need to ascertain changes in wealth. In some instances, it may be easier to account for changes in wealth than to estimate the total value of wealth. Changes in wealth entail gross

investments (in physical and human capital) minus depreciation and depletion (of physical, human, and natural capital.)

36. Although the information about some central aspects of household wealth is in principle available from national accounts balance sheets, information is often incomplete. Furthermore, certain assets are not recognized as such in the standard accounting framework. A particular important one is human capital. Those studies that computed human capital stocks found that they account for an overwhelming part (80% and more) of all wealth¹. A systematic measurement of human capital stocks is of interest from a number of perspectives. It constitutes an integral part of an extended measure of household production (see below), and it is an input for the construction of sustainability indicators.

37. Note a fundamental problem with valuing stocks. When there are markets for assets, the prices at which assets are bought and sold serve to value the stock as a whole. There may be no markets or no trading on markets as has recently been the case for certain financial assets and the question arises how to value them. Even when there are market prices, they only correspond to the small fraction of the stock that is really transacted and they may be so volatile as to put a question mark on the interpretability of balance sheets. That said, basic information on assets and liabilities is key to assessing the economic health of sectors and to the financial risks to which they are exposed.

¹ See, for example, Jorgenson and Fraumeni (1989).

4. Bringing out the household perspective

38. Income can be computed for the economy as a whole but also for private households. Some of the income of citizens is taken away in the form of taxes. This is money that is not at their disposal. But the government takes this money away for a reason: to provide public goods and services, to invest for example in infrastructure and to transfer income to other (normally more needy) individuals. Thus other households receive transfers from government, and all households receive benefits. A commonly employed measure adds and subtracts these transfer payments. The resulting measure is referred to as a measure of *disposable income for households*. However, disposable income only captures monetary transfers between households and the government, thereby neglecting the in-kind services that government provides.

4.1. Adjusting household income measures for government services in kind

39. Earlier in this text we mentioned the invariance principle according to which *a movement of an activity from the public to the private sector, or vice versa, should not change our measure of performance, except to the extent that there is an effect in quality or access*. This is where a purely market-based measure of income or economic performance meets its limits and where a measure that corrects for differences in institutional set-up may be warranted for comparisons over time or across countries. *Adjusted disposable income* is a national accounts measure that goes some way towards accommodating the invariance principle, at least where ‘social transfers in kind’ by government are concerned.

40. The meaning of adjusted disposable income is best explained by way of an example. Assume that labour income in society equals 100 and that individuals that are active in the

labour market buy private health insurance. They make an annual payment for the insurance equal to 10, which can be decomposed into 8 units of insurance premiums (an actuarial value of losses of 8) and 2 units of consumption of insurance services. At the same time, persons who are sick receive 8 units as a reimbursement of their health expenditures. This is Case A in the table below: no taxes are paid and insurance claims and premiums just offset each other so that disposable income for households equals 100. Now, assume that the government decides to provide the same amount of health insurance coverage to everyone, funded through a tax of 10 units. Nothing has changed, other than that the government is now collecting the insurance payment and distributing the benefits (Case B). But according to standard national accounts statistics, disposable income has fallen, to 90 currency units. Thus, disposable income yields a distorted comparison. If one adds in the social transfers in kind that households receive from the government under Case B (8 units corresponding to the reimbursement of health expenditures and 2 units corresponding to the running costs of the insurance), the adjusted measure of household disposable income indicates equality between the two cases.

41. The above example leaves, however, aside any consideration about which insurance regime operates more cost effectively and about profits that can be made by private insurance companies— it was simply assumed that private and public insurance services are equivalent to 2 currency units. In practice, this is almost certainly not the case although it is difficult to establish a general observation on relative efficiency of schemes. If the financial services industry is not perfectly competitive (a reasonable assumption in most countries), the transfer of responsibility from the private to the public sector will be reflected in decreased profits, decreased value-added and decreased income. But the profits are simply a transfer payment from households to firms. Society is not better off as a result of these profits.

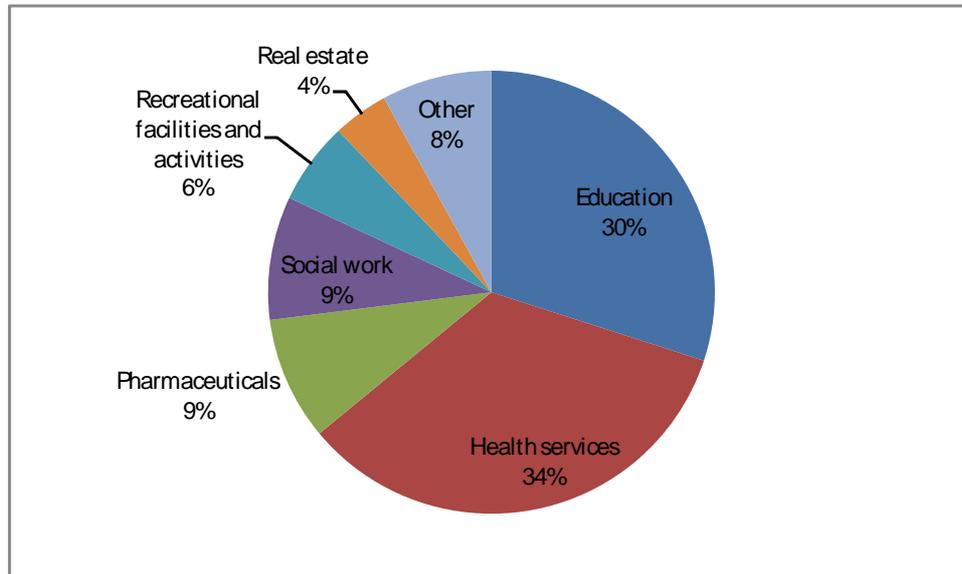
	Private insurance scheme (Case A)	Public insurance scheme (Case B)
Labour income	100	100
Tax	0	-10
Insurance premiums (excluding insurance services)	-8	0
Insurance claims	+8	0
Household disposable income	100	90
Social transfers in kind : <i>reimbursements</i>	0	+10
<i>running costs of the insurance</i>		+8
		+2
Adjusted household disposable income	100	100

42. While the failure to estimate the value of security services provided causes one set of biases, there are other biases that arise from the fact that the value of some social transfers in kind (those corresponding to the running costs of the insurance in the example above) is measured by the costs of producing these services. In some countries, in particular in the developing world, the cost of these services may greatly exceed their value to households, who may receive little or nothing. The result is large scale overestimation of the level of household income and consumption. Some of this can be tackled by using output-based volume measures for health and education services produced by government. And it is likely that different parts of the population benefit differently from social transfers in kind provided by government. This is an important distributional aspect.

43. Major items in social transfers in kind are health and education services, subsidized housing, sport and recreation facilities and the like that are provided to citizens at a low price or for free. In France, general government provides nearly all of these services at a cost of about €290 billion in 2007. Education and health services each account for about

1/3 of total transfers in kind, and housing and recreational and cultural activities (museums, public parks...) account for about 10% as indicated in Figure 4.

Figure 4 Social transfers in kind from general government, France 2007



Source: INSEE.

4.2. Medians and means - distribution of market income

44. Average measures of income per person are helpful but give no indication about how available resources are distributed across persons or households. For example, *average* income per capita can remain unchanged while the distribution of income becomes less equal. It is therefore necessary to look at disposable income information for different income groups. A conceptually simple way of capturing distribution aspects is to measure *median* income, the income such that half of all individuals are above that income, half below. The median individual is, in some sense, the “typical” individual. With increasing inequality, there may be increasing differences between median and average income; a focus on average income does not give an accurate picture of the economic well-being of the ‘typical’ member of society. If all the increases in societal income accrue, say, to those in the top 10%, median income may remain unchanged, while average income increases.

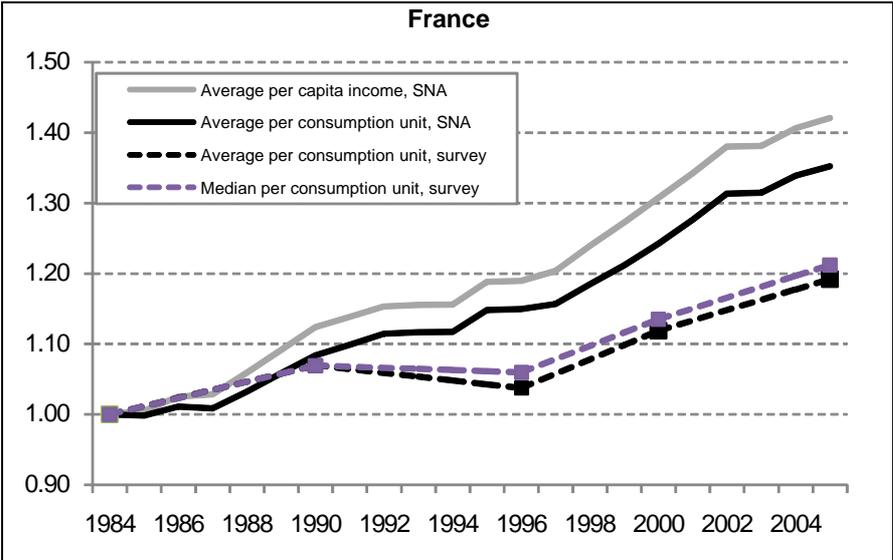
Over the past two decades, the dominant pattern in OECD countries is one of a fairly widespread increase in inequality, with strong rises in Finland, Norway, Sweden (from a low base) and Germany, Italy, New Zealand, and the United States (from a high base). In these cases, the two numbers would give different pictures of what is happening to societal well-being. Alternatively, the development of disposable income of different income groups can be followed. Such an approach would, for instance, look at the numbers below a critical poverty level, the average income of those in the bottom or top decile.

45. In practice, moving from average to median income is more difficult than meets the eye. Measures of average income are obtained by dividing total income by a population figure. To consider distributional elements, micro-economic information is needed that provides income information for individual households or groups of households. Micro-economic measures of household income refer to people living in private households and are typically derived from household income surveys whereas macro-economic measures are provided by the national accounts.

46. An important choice lies also with the unit of measurement. Macro-estimates give totals for a whole country or sector, while micro data retain the household (or the family) as unit within which resources are pooled and shared, and adjust income for differences in “needs.” There are, for instance, fixed costs to running a household, allowing larger families with the same per capita income to have a higher standard of living. Another step towards bringing demography and some distributional aspects into income measures is to calculate disposable income per consumption unit or per household rather than per person. Consumption units are households with an adjustment for their size so that account is taken of the economies of scale in housing and other costs. This adjustment is of increasing importance as the size of households changes.

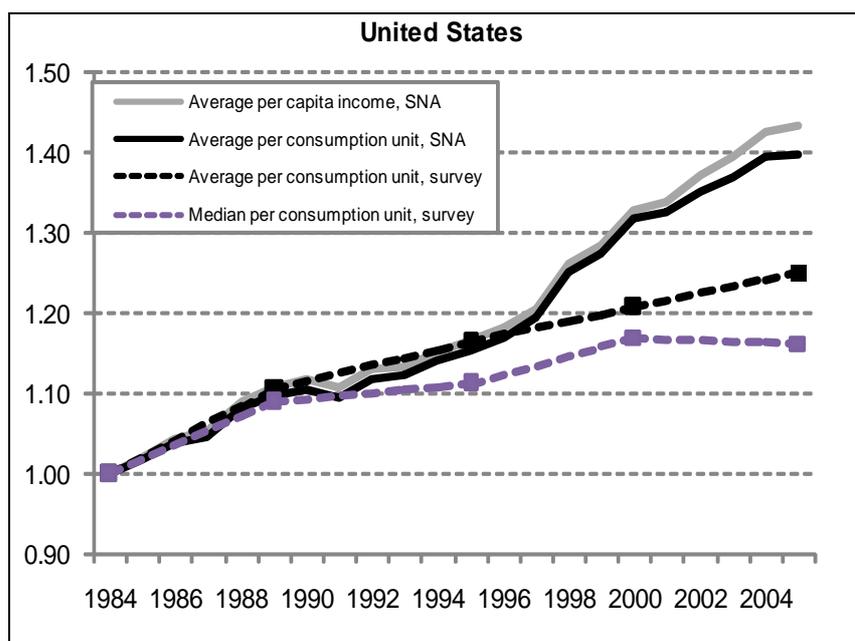
47. Against this background, we can consider the evolution of average and median household income in several countries. Figure 5 and Figure 6 show results for France and for the United States. Average income per capita and average income per consumption unit diverge, reflecting a trend towards smaller household size. Survey income measures permit comparing average and median income. In the case of France, these two items move in parallel. At least from this perspective, there is thus no indication of a widening income distribution. The picture is different for the United States where average income per capita and per consumption unit grow at the same rate but where there is widening gap between median and average income, pointing to a more unequal income distribution.

Figure 5: Trends in different measures of household disposable income



Source: Computations based on OECD SNA and income distribution data.

Figure 6: Trends in different measures of household disposable income



Source: Computations based on OECD SNA and income distribution data.

48. There are many measurement issues that can influence the above statements. An important source of discrepancy between micro and macro estimates is property income, whether imputed or not. If this aggregate is not well measured in micro estimates, this could explain why average and median incomes in these estimates have a parallel evolution in France where wage inequalities are less important than property income inequalities. In addition, there is a possibility that top incomes are under-represented in household income surveys. Finally, the international comparability between household surveys is far from perfect.

49. From a perspective of living standards what matters is that the distribution of income and wealth determines who enjoys access to the goods and services produced within a society. Complementing measures of average income by income measures with a distributional element is thus a crucial task for official statistics. Ideally, such

distributional measures are compatible in scope with average measures from the national accounts.

50. Income distribution should not only be judged by the distribution of nominal income. The same dollar may buy different bundles of product, depending on the income group of the purchaser. Going from nominal to real income means applying a price index and a point of relevance is therefore *'whose price index can we measure?'* Often, conceptual discussions about price indices are conducted as if there were a single representative consumer. Statistical agencies calculate the increase in prices by looking at what it costs to purchase an average bundle of goods. The problem is that different people buy different bundles of goods. Poor people spend more on food; rich on entertainment. They also buy goods and services in different types of stores, which sell "similar" products at very different prices. When all prices move together, having different indices for different people may not make much a difference. But recently, with soaring oil and food prices, these differences have become marked. Those at the bottom may have seen real incomes be more affected than those at the top.

51. A *price index* for (actual) private consumption *for major groups in society* (age, income, rural/urban) *is necessary* if we are to appraise their economic situation. One of the recommendations of the *Commission sur la mesure du pouvoir d'achat des ménages* (2008) (Commission on the measurement of purchasing power of households) in France was to develop consumer price indices for owners of dwellings, for households who rent dwellings and for households that are about to purchase dwellings. A full development of price indices differentiated by socio-economic groups requires, however, that different prices have to be collected for different parts of the population, so that socio-economic aspects are taken into account in data collection design. This is likely to turn out difficult and costly. Its development should constitute a medium-term research objective – a

recommendation that echos a similar conclusion by the 2002 *Panel on Conceptual, Measurement, and other Statistical Issues in Developing Cost-of-Living Indexes* in the United States. Such work would not only foster the quality of deflation procedures, it would also make it easier for citizens to identify their personal situation with some of the income and price data released by statistical offices.

4.3. Broader measures of households' economic activity

52. There have been changes in the functioning of households and the society. For example, in the past, more people received services from their family that are now provided purchased on the market. This shift translates into a rise of income as measured in the national account and gives a false impression of a change in living standards, while it only reflects a shift from non-market to market provision of services. Just as we argued that a basic principle in national accounts measurement was that a shift from private to public provision of a particular good or service or vice versa should not affect measured output, so too, a shift of production from market to household production, or vice versa, should not affect measured output. We noticed earlier that in practice, current conventions do lead to changes in measured income in both instances.

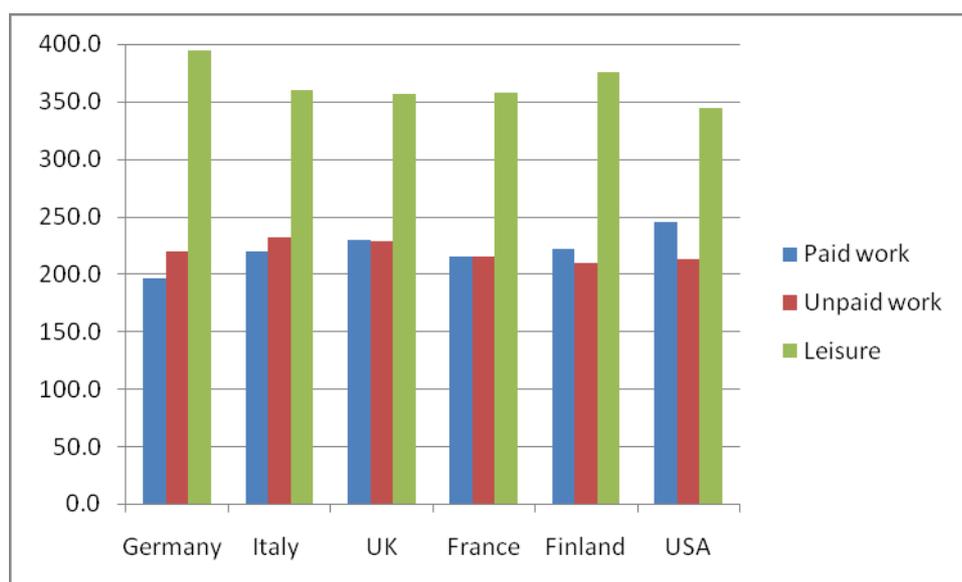
53. Imagine a two-parent household with two children with an income of \$50,000 a year, in which only one parent works full-time for pay and the other specializes in home production. The parent who stays at home does all shopping, cooks all meals, does all cleaning, and does all child care. As a result, this household does not need to devote any of its market income to such purchases. Now, imagine a two-parent household with two children in which both parents work full-time for pay, and neither parent has any time left over for household production or child care. They must pay for shopping, cooking, cleaning and child care out of pocket. Their available income is reduced. Standard

measures of family living standards treat these households as if they have identical living standards but obviously they don't. Focusing on market production only provides a biased picture of living standards – some of the measured increase in market production may simply reflect a *shift* of the locus of production from households to the market.

54. To get a sense for the economic importance of home production, one has to start by examining how persons use their time. Figure 7 provides a first comparison of time spent per household and day on various activities. Household production comprises time spent on housework, purchasing goods and services, caring for and helping household and non-household members, volunteer activities, telephone calls, mails and email and travel time related to all these activities. 'Personal care' is mainly sleeping, eating and drinking, whereas leisure was defined to include, sports, religious and spiritual activities and other leisure activities.

55. Against those caveats, one finds that more time is spent on household production in European countries as opposed to the United States. More time is spent on leisure in Finland, France, Italy, Germany and the United Kingdom than in the United States (Figure 7). Note that some of the classifications are ambiguous and results should therefore be read with care. For example, eating and drinking are included in the definition of personal care whereas arguably, part of eating and drinking is time spent on leisure. Many view cooking — and then eating - a good meal as a most enjoyable leisure activity, not a chore that is easily substitutable with a meal in a fast food restaurant. And the time use picture can change if eating time is allocated differently. We conclude that the allocation of specific activities to time use categories as well as their international comparison leaves room for improvement and harmonisation.

Figure 7: Housework, paid work and leisure
Minutes per day and person, latest year available*



*using normalised series for personal care; Unites States: 2005, Finland 1998, France 1999, Germany 2002, Italy 2003, United Kingdom 2001.

Source: OECD (2009), based on HETUS and ATUS databases.

56. A major gap in time use data is availability of consistent time series. Time use surveys have been conducted in the past but, in most cases, not periodically, and often comparability between surveys is limited. The assessment of time use over longer periods has to take recourse to approximations and estimates of varying quality. This also holds for the present study. An important task for the future will be to build up nationally and internationally comparable time series of information on how people use time. Such work is under way in the United States and in several European countries but lacking in many other parts of the world.

57. It is now possible to come up with an illustrative calculation for the value of household production for France, Finland and for the United States. The approach chosen here is simple: the value of the production of household services is measured by its costs. The value of labor is estimated by applying a wage rate of a generalist household worker to the number of hours that persons spend on housework. Methodology matters in this

context and results can differ markedly, depending in particular on the hypotheses chosen for the valuation of labor and capital. We are also lacking estimates for productivity change in household production.

58. However, our estimates provide orders of magnitude. It is apparent, and no surprise in light of previous studies that imputations for own-account production of household services are a sizeable matter. Our calculations show that household production amounts to about 35% of conventionally-measured GDP in France (average 1995-2006) and to about 40% in Finland and 30% in the United States over the same period.

59. Once one starts thinking about non-market income, one has to think about leisure. With time spent on generating income (market or non-market) we buy or produce goods and services to meet our needs or for simple enjoyment. Time available for leisure obviously affects well-being. Changes in the amount of leisure over time and differences between countries represent one of the more important aspects of changes in well-being over time and differences across countries. Focusing only on goods and services can therefore bias comparative measures of living standards towards the production of goods and services. This is of particular concern as the world begins to come to terms with environmental constraints. It will not be possible to increase the production especially of goods beyond limit. Taxes and regulations will be imposed that will discourage the production of goods. It would be a mistake if, as a result, leisure time increases, that we judge living standards to have decreased. As society progresses, it is not unreasonable to expect people to enjoy some of the fruits of that progress to be taken in the form of leisure. Different societies may respond, however, differently, and we do not want to bias our judgments (e.g. of success) *against* societies that choose to enjoy leisure.

60. Measurement of the value of leisure starts again from time use data. We multiply the average leisure time per day by the working age population and by the average wage rate in the economy. Again, there are many measurement issues associated with this procedure but the purpose here is to show that estimates are feasible and can produce meaningful results, including for cross-country comparisons. For the three countries at hand, the value of leisure about doubles net disposable income of households in nominal terms. More interesting than nominal income levels is the question how considering leisure affects the measured growth rates of *real* income and their comparisons across countries. This is captured in Table 2. It shows the evolution of household income, now adjusted for household work (upper panel) and for household work and leisure (lower panel). For all countries, the new real income measures grow more slowly than the traditional measures of income. When expressed as income per consumption unit (i.e., per household, adjusted for household size), income growth rates between the three countries turn out to be very similar.

Table 2: Household income in real terms
Percentage change at annual rate, 1995-2006

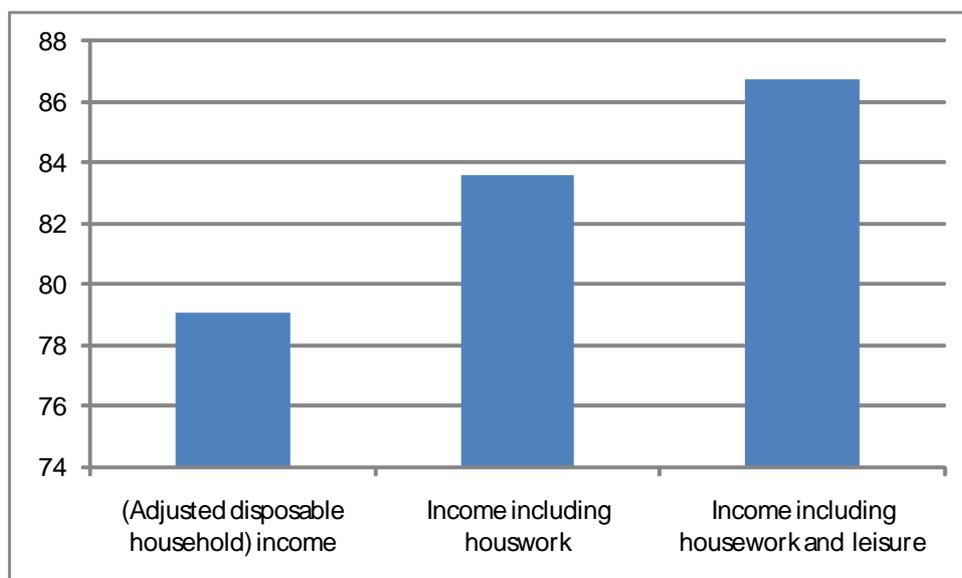
	France	United States	Finland
Adjusted disposable income plus housework			
Total	1.9%	2.9%	2.0%
Per consumption unit	1.1%	1.7%	1.6%
Adjusted disposable income plus housework and leisure			
Total	1.4%	2.3%	1.4%
Per consumption unit	0.7%	1.0%	0.9%

61. The imprecision associated with the above estimates is reiterated here. These are orders of magnitude at best and should not be over-interpreted. However, it is clear that recognition of broader measures of economic activity and recognition of leisure do make a

difference to comparisons over time and between countries. More work needs to be done to test methodologies, to single out the most critical parameters and to test robustness of such measures. Only if there is sufficient confidence in extended measures of income, will there be a broader take-up by statistical offices.

62. Even more instructive than estimating the rate of change in real income is to assess how household production and leisure bear on the comparison of income *levels* across countries. Income levels should be compared in real terms and we construct currency converters, so called *Purchasing Power Parities (PPP)* that permit comparisons of real ‘full’ income (including housework and leisure) across countries. Figure 8 compares three income aggregates between France and the United States. The first comparison is with the established adjusted disposable income measure. Here, France’s per capita income is just below 80% of the comparable United States figure. Adding in the income equivalent of unpaid housework narrows the gap to 84%. If in addition, leisure is accounted for, one ends up with a relative income level of 87%.

Figure 8: Real income per capita in France compared to the United States, 2005
Unites States = 100



4.4. Distribution of full income

63. It was argued earlier that measures of average income should be accompanied by income measures with distributional information such as median disposable income, the income such that half of all individuals are above that income, half below. The rationale for examining income distribution holds for market income but also for broader measures of income such as full income. The recognition of own-account production of household services and leisure may not only affect aggregate measures of income and production. It also changes the established picture of income distributions, as confirmed by the literature on this topic.
64. Developing distributional measures of full income is, however, a formidable task. The single most difficult item is to allocate to income groups those income flows that have been imputed at the macro level when comprehensive measures of income were derived. For example, imputed rents from own-occupied housing constitute such an item. Other imputations for own-account services produced by households also fall under this category. And so do the distributional effects of government services that are provided in kind.
65. Again, measurement difficulties should not prevent us from pursuing the objective of getting a more comprehensive picture of the distribution of income and wealth.

CHAPTER 2 – QUALITY OF LIFE

1. Introduction

66. Quality of life is a broader concept than economic production and living standards. It includes the full range of factors that makes life worth living, reaching beyond its material side. While some extensions of economic accounting allow including some of the elements that shape quality of life in conventional measures of economic well-being, all approaches based on resources (or on people's command over commodities) remain limited in important ways. First, many resources are not marketed: even when they are, prices will differ across individuals, making it problematic to compare real income across people. Second, many of the determinants of human well-being are not resources but aspects of people's life-circumstances: they cannot be described as resources with imputable prices, even if people do make trade-offs among them. Lastly, resources are means that are transformed into well-being in ways that differ across people: those individuals with greater capacities for enjoyment or greater abilities for achievement in valuable domains of life may be better off even if they command fewer economic resources. These arguments, by themselves, suggest that resources are an inadequate metric for quality of life. Which other metric should be used for assessing quality of life depends on the philosophical perspective taken.

67. While a long tradition of philosophical thought has addressed the issues of what gives life its quality, recent advances in research have led to new and credible measures. These measures, while not *replacing* conventional economic indicators, provide the opportunity to *enrich* policy discussions and to inform people's view of the conditions of the community where they live. More importantly, they have now the potential to move from research to standard statistical practice. While some of these measures reflect

structural conditions that are relatively invariant over time but that differ systematically across countries, others are more responsive to policies and suitable for monitoring changes over short-periods of time. Both types of indicators play important roles in measuring quality of life.

2. Conceptual approaches to measuring quality of life

68. Three conceptual approaches have retained the attention of the Commission as useful in thinking about the measurement of quality of life.

69. The first approach, developed in close connection with psychological research, is based on the notion of *subjective well-being*. A long philosophical tradition views individuals as the best judges of their own conditions. This approach is closely linked to the utilitarian tradition but has a broader appeal due to the strong presumption, in many streams of ancient and modern culture, that enabling people to be ‘happy’ and ‘satisfied’ with their life are universal goals of human existence. Research on subjective well-being holds the promise of delivering not just a good measure of the level of quality of life but also an understanding of its determinants.

70. The second approach is rooted in the notion of *capabilities*. This approach conceives a person’s life as a combination of various ‘doings and beings’ (functionings) and of his or her freedom to choose among these functionings (capabilities). Some of these capabilities may be quite elementary, such as being adequately nourished and escaping premature mortality, while others may be more complex, such as having the literacy required to actively participate in political life. The capability approach, which has strong roots in philosophical notions of social justice, rests on a number of foundations. These include a

focus on human ends and on respecting people's ability to pursue and realise goals that he or she values; a rejection of the economic model of individuals acting to maximise their self-interest devoid of relationships and emotions; an emphasis on the complementarities between various capabilities; and a recognition of people's diversity, which brings attention to the role played by ethical principles in the design of the 'good' society.

71. The third approach, developed within the economics tradition, is based on the notion of *fair allocations*. The basic idea, which is common to welfare economics, is that of weighting the various non-monetary dimensions of quality of life (beyond the goods and services that are traded in markets) in a way that respects people's preferences. This approach requires choosing a particular reference point for each of the various non-monetary dimensions, as well as information on people's current situations and on their preferences with respect to these points. This approach avoids the pitfalls of basing evaluations on an 'average' willingness-to-pay that may disproportionately reflect the preferences of those who are better-off in society by focusing on equality among all of its members.

72. These approaches have obvious differences, but also similarities. For example, subjective well-being is sometimes claimed to encompass all capabilities, to the extent that these are attributes and freedoms that people *value* (implying that their achievements will improve people's subjective states). However, proponents of the capability approach also emphasise that subjective states are not *all* that matter, and that expanding people's opportunities is important in itself, even if this does not show up in higher subjective well-being. Similarly, both the capability and the fair allocation approaches rely on information on the objective attributes of each person, while differing in the ways in which these are weighted and aggregated. While the choice between these approaches is ultimately a normative decision, all these approaches point to the importance of a number of features

for quality of life that go beyond command over resources. Measuring these features requires the use of types of data (i.e. responses to questionnaires and non-market observations of personal states) that are not captured by market transactions.

3. Subjective measures of quality of life

73. For a long time, economists have assumed that it was sufficient to look at people choices to derive information about their well-being, and that these choices would conform to a standard set of assumptions. In recent years, however, much research has focused on what people value and how they act in real life; this research has highlighted large discrepancies between standard assumptions of economic theory and real world phenomena. A significant part of this research has been undertaken by psychologists and economists based on subjective data on people's reported or experienced well-being.
74. Subjective measures have always been part of the traditional tool-kit of economists and statisticians, as many features of our economies and societies (e.g. unemployment) are measured through people's responses to a standard set of questions. The specific feature of the subjective measures of quality of life discussed here is that what people report about their own conditions has no obvious objective counterpart: we can compare 'perceived' and 'actual' inflation, for example, but only respondents can provide information on their own subjective states and values. Despite this feature, a rich literature on these subjective measures concludes that they help predict people's behaviour (e.g. workers who report more dissatisfaction in their work are more likely to quit), and that they have validity with respect to a range of other information (e.g. people who report themselves as 'happy' tend to smile more and to be rated as happy by people around them; these self-reports are also correlated with electrical readings of the brain).

75. Subjective approaches distinguish between the *dimensions* of quality of life (i.e. people's subjective experiences and their evaluations) and the objective *factors* shaping them. In turn, subjective dimensions of quality of life encompass several aspects. The first is represented by people's evaluations, either of their life as a whole or of its various domains, such as family, work and financial conditions. These evaluations imply a cognitive exercise from each person, and an effort to take stock and summarise the full range of elements that people value (e.g. their sense of purpose, the fulfilment of their goals and the regards of others on one-self). The second aspect is represented by people's actual feelings, such as pain, worry and anger, or pleasure, pride and respect. To the extent that these feelings are reported in real time, they are less affected by biases due to memory and by social pressures related to what is deemed to be 'good' in society. Within this broad category of people's feelings, the research on subjective well-being distinguishes between positive and negative affects, as both may characterise the experience of each person.
76. All these aspects of subjective well-being (cognitive evaluations, positive affects and negative affects) should be measured separately to get a satisfactory appreciation of people's life. Which of these aspects matter more, and for what purpose, is still an open question. Much evidence suggests that people act to achieve satisfaction in their choices, and that choices are based on memories and evaluations. But memories and evaluations can also lead to systematic errors while some choices are made unconsciously rather than weighting the pros and cons of various alternatives.
77. Subjective measures of people's life-evaluations and affects provide measures of quality of life that can be monitored over time; some of these measures can also be compared across countries in reliable ways. Probably more importantly, however, is that these measures inform about the determinants of quality of life at the level of *each person*.

These determinants include both individual conditions and features of the environment where people live. These determinants also vary depending on the aspect considered. For example, people's activities (such as commuting, work, or socialising) are more important for affects, while people's conditions (such as being married, or having a rewarding job) are more important for life evaluations. In both cases, however, these measures provide information beyond that conveyed by income. For example, younger and older people in most developed countries report higher evaluations of their life than prime age people, a pattern that contrast sharply with that for income.

78. One point where various subjective measures of people's well-being agree is in pointing to the large costs of unemployment for people's quality of life. People who become unemployed report lower life-evaluations, even after controlling for their lower income, and with little adaptation over time; unemployed people also report higher prevalence of various negative affects (sadness, stress and pain) and lower levels of positive ones (joy). These subjective measures suggest that the costs of unemployment exceed the income-loss suffered by those who lose their jobs, reflecting the existence of non-pecuniary effects among the unemployed, and of fears and anxieties generated by unemployment in the rest of society.

79. While important advances in the measurement of subjective well-being have been realised following initiatives of individual researchers and commercial data providers, these data remain limited in terms of the statistical inferences that they allow. National statistical systems need to build on these efforts, and incorporate questions on various aspects of subjective well-being in their standard surveys. They should also develop longitudinal studies that could support more valid inferences on the relative importance of the various factors at work.

4. Objective features shaping quality of life

80. Both the capability and the fair allocation approaches give prominence to the objective conditions and opportunities available to people, while differing in how these features are valued and ranked. While these objective features may also have an instrumental value for subjective well-being, both of these conceptual approaches regard an expansion of people's opportunities in these domains as intrinsically important for people's life.

81. The range of objective features to be considered in any assessment of quality of life will depend on the purpose of the exercise: whether the goal is to assess changes in conditions within national jurisdictions, or to compare these conditions across countries at different levels of development. Some may matter as descriptors of people's states (e.g. health), while others may inform about the freedoms that people have to pursue the goals that they value (e.g. political voice). In general, all these measures highlight that how societies are organised makes a difference for people's quality of life, and that their influences are not all captured by conventional measures of economic resources.

4.1. Health

82. *Health* is a basic feature shaping both the length and the quality of people's life. Its assessment requires good measures of both mortality and morbidity. Data gaps remain significant in both fields. Mortality statistics by age and gender document the risk of death confronting people and are used to calculate the expected length of a person's life. These indicators are today available in all developed countries but remain limited in large parts of the developing world, in particular for adults, and this limits the possibility of

monitoring progress in achieving the UN *Millennium Development Goals*. Further, age-specific mortality statistics are vectors: to obtain a scalar measure of people's lifespan, they need to be aggregated in suitable ways and standardised for differences in age-structure across countries and over time. While different aggregation formulas and standardisation methods exist, they lead to different numerical results and rankings when comparing countries with different balances between child and old-age mortality and with survival curves that cross each other. This suggests that a variety of mortality measures should be compiled and regularly monitored.

83. The state of advancement is far more limited for statistics on morbidity, a situation that has led to long-standing disagreements on whether or not declines in mortality have been matched by similar declines in morbidity. Existing measures of morbidity rest on a variety of sources: records of people's height and weight; diagnostics from health professionals; registers for specific diseases; and self-reports drawn for censuses and surveys. Some of these measures relate to the occurrence of diseases or injuries, while others refer to their consequences in terms of functionings of the person affected (which also depend on the quality of treatment). The variety of measures and underlying data is inevitable given the many manifestations of poor health, but is also a real obstacle to comparing countries and monitoring changes in people's health status over time. Measures are even sparser when moving from physical to mental disorders, despite evidence that these affect (at least in mild forms) a large share of people, that most of these disorders go untreated, and that their incidence has been increasing in some countries.

84. The variety of dimensions of people's health has led to several attempts to define a summary measure that combines both mortality and morbidity. Several combined indices of people's health exist, but none currently commands universal agreement. Further, they

all inevitably rest on ethical judgements that are controversial, and on weights for various medical conditions whose legitimacy is not always clear.

85. The challenges posed by this variety of health measures are not confined to cross-country comparisons but extend to within-country comparisons. Recent research on inequalities in health status has highlighted several patterns. First, people with lower education, lower income and from lower occupational classes tend to die at younger ages and to have, within their shorter life, a higher prevalence of various health problems. Second, these differences do not reflect only the worse outcomes for people at the very bottom of socio-economic scale but are observed throughout the distribution. Last, there is no consistent narrowing of these between-groups health inequalities over time, which have rather increased in some countries. While these patterns have an obvious relevance for assessing quality of life, existing measures do not allow cross-country comparisons of the size of these inequalities due to differences in the measures of health outcomes used, in the individuals' characteristics considered (education, income, ethnicity), and in the reference population and geographic coverage of the various studies.

4.2. Education

86. A long tradition of economic research has stressed the importance of education in providing the skills and competencies that underpin economic production. But education matters for quality of life independently of its effects on people's earnings and productivity. Education is strongly associated with people's life-evaluations (based on the ladder-of-life scale), even after controlling for their higher income. Further, more educated people typically have better health status, lower unemployment, more social connections, and greater engagement in civic and political life. While available evidence does not always allow concluding on the direction of causation between education and

these other dimensions of quality of life (e.g. less healthy children may miss school more often), the consensus is that education brings a range of returns (monetary and non-monetary) that benefit both the person investing in education and the community in which they live. Measuring the size of these wider benefits of education is an important research priority, but progress requires better measures of people characteristics in a range of domains and to follow the same individual over time.

87. Available educational indicators cover a broad range of fields. Some refer to inputs (e.g. school enrolment, educational expenditures and school resources), others refer to throughputs and outputs (e.g. graduation rates, number of completed years of schooling, standardized test measures of people's achievements in terms of literacy and numeracy). Which of these indicators is more relevant depends on the stage of development of each country and on the goal of the evaluation exercise. The available indicators highlight large differences across countries, with various educational indicators sometimes highlighting contrasting patterns. Within countries, measures of inequalities in learning outcomes are especially important for youths at the bottom of the achievement scale and who are at risks of poverty or exclusion from well-paid and rewarding jobs in adult life. As education is an important predictor of many dimensions of people's life, all social surveys should systematically include information on the learning experiences of respondents and of their parents, as well as information on other features shaping the quality of their life.

88. Some of the most relevant indicators for assessing the role of education for quality of life are measures of people's competencies, and several tools have been developed in recent years to measure them in standardised ways. These measures, however, remain limited in important ways. First, and most obvious, not all countries currently implement these surveys. Second, many of these tools were not developed from the perspective of measuring people's capabilities in a broad sense but for the purpose of assessing

educational policies, which typically required focusing on a more narrow set of measurable competencies. Third, existing assessment tools often have a narrow coverage, as schooling is only one of the inputs that lead to knowledge, skill development and improvements in quality of life. Information on the experiences and ‘soft’ competences learned by children in their early years remain limited despite increasing evidence that early-childhood experiences matter for people’s learning and quality of life in later years; measurement tools remain limited also for comparing competences of students in higher education and for assessing workers’ experiences in terms of adult education and training, although this will change as new surveys of adult competencies are developed and implemented. As for other features of quality of life, however, the main problem for indicators in this domain is not the lack of detailed information on education *per se*, but rather the lack of surveys measuring both education and other outcomes that matter for quality of life at the individual level.

4.3. Personal activities

89. How people spend their time, and the features of people’s personal activities, matter for quality of life irrespectively of the income that they generate. The activities in which people spend their time have effects on people’s subjective well-being, both in terms of their hedonic experiences and of their evaluative judgements. More generally, people cannot always ‘choose’ among these various activities in the same way as they allocate their budget among various goods and services, due to a lack of effective alternatives; further, these choices will generally affect other people within the family and community, and some of these personal activities effectively represent indirect costs to production (e.g. commuting) rather than consumption. Because of both political demands and the feasibility of providing concrete and comparable measures, the main activities discussed

here are paid work, commuting, unpaid work and leisure time. Housing, although not representing an activity *per se*, is also discussed because it is the setting for a number of personal activities and because of its importance for the quality of life.

90. Paid work matters for quality of life because it provides identity to people and opportunities to socialise with others. Not all jobs are equally valuable in this respect, and this underscores the importance of collecting more systematic information on the quality of paid work, as done in the context of ongoing work on ‘decent work’ pursued by a number of international organisations. Some national surveys provide information on many aspects of decent work such as non-standard employment, gender gaps in employment and wages, discriminations in the workplace, opportunities for lifelong learning, access to employment for disabled persons, working time and ‘unsocial hours’, perceived work-life balance, work accidents and physical risks, work intensity, work-related health problems, social dialogue and workers’ autonomy. Their practical use is however limited by their small sample size and by differences across countries.
91. Commuting time is also a key feature of the quality of work, and its monitoring requires information on the number of hours spent travelling to and from work during a specified period, as well as on accessibility and affordability of transports.
92. Unpaid domestic work, such as shopping, care of children and other household members, is important from the perspective of assessing both the total amount of household services produced and how family chores are distributed between men and women.
93. A long tradition of research has emphasised the importance of leisure-time for quality of life. This points to the importance of developing indicators of both its quantity (number of hours) and quality (number of episodes, where they took place, presence of

other people), as well as measures of participation in cultural events and of ‘poor leisure’ (such as the share of children who did not take a holiday away from home in the previous year).

94. Finally, despite the importance of housing for a variety of social outcomes (such as children education), no core set of housing indicators currently exists for international comparisons: remedying this situation would require better information on the number of people who are homeless or living in emergency shelter, and on housing quality (e.g. in terms of the environmental services available and overcrowding).

95. In several cases, suitable indicators in these various fields already exist, and the challenge is that of improving upon what has been achieved in the past. In other areas, however, existing measures remain seriously deficient, and progress requires investment in new statistical capacity. A case in point, cutting across all the personal activities described above, is that of measuring how people spend their time. Time is the natural metric for comparing personal activities, and one priority should be to develop measurement tools grounded on clear definitions and based on surveys with a consistent design, representative of patterns over a full year, and undertaken with sufficient regularity (all requirements that are not often met). These surveys should inform on both the amount of time spent in various activities and on the enjoyment they provide. This is important as the same activity will generate different hedonic experiences depending on people’s own conditions (e.g. whether they are unemployed or not); this information also matters for assessing inequalities among different groups in society. While these investments in statistical capacity are costly, and compete with other priorities, their pay-off is potentially huge.

4.4. Political voice and governance

96. Political voice is an integral dimension of the quality of life. Intrinsicly, the ability to participate as full citizens, to have a say in the framing of policies, to dissent without fear and to speak up against wrong are essential freedoms and capabilities. Instrumentally, political voice can provide a corrective to public policy, ensure accountability of officials and public institutions, reveal what people need and value, and call attention to significant deprivations. Political voice also reduces the potential for conflict and enhances the prospect of building consensus on key issues, with pay-offs for economic efficiency, social equity, and inclusiveness in public life.

97. The opportunities for expression and the degree of responsiveness of the political system depend on the institutional features of each country, such as the presence of a functioning democracy, universal suffrage, free media, and civil society organisations. It also depends on some key aspects of governance, such as legislative guarantees and the rule of law. Legislative guarantees include both constitutional rights, and rights provided by general laws that enhance the quality of life of all residents and that reflect the social consensus prevailing in different countries and times. The structure of laws can also affect the investment climate in a country and thus have an impact on market functioning, economic growth, job creation and material welfare. Further, to realise their potential, legal guarantees require effective implementation and substantive justice, which depends on how various institutions (e.g. the police, the judiciary and various administrative services) function, whether they are free from corruption, political interference and social prejudice, and whether they can be held accountable for their decisions.

98. Comparisons based on existing indicators of political voice, legislative guarantees and the rule of law highlight vast differences between countries, especially between those

with a long history of democratic functioning and those that have moved from authoritarian to democratic regimes only more recently and that have not yet established the full range of freedom and rights. Even in the developed world, however, low trust in public institutions and declining political participation point to a growing gap in how citizens and political elites perceive the functioning of democratic institutions. The exercise of political voice, fundamental rights and opportunities for civic participation in these countries also differs systematically across groups of people, especially between citizens and the large numbers of international migrants.

99. Indicators of political voice and democratic governance should help evaluate the functioning of multiparty democracy and universal suffrage, the degree of decentralisation in government decisions, the presence of a free media and various freedoms (e.g. to form and join civil organizations, trade unions and professional bodies, or to participate in civic and social activities). Relevant indicators should cover the rights embedded in constitutions, laws (e.g. those that promote civil and criminal justice, equality, inclusion, accountability and affirmative action), international covenants on human rights and basic freedoms, as well as the functioning of the judicial system (e.g. its independence from corruption and political influences, the speed with which it delivers justice, and its accessibility to both citizens and residents). Indicators of many of these aspects are typically compiled by bodies outside the boundaries of national statistical systems and are mainly based on the opinion of experts. These indicators need to be complemented, and in some cases replaced, by surveys on citizens' own perceptions about the functioning of political, legal and executive institutions, the difficulties they face in accessing them, and the trust that they place in them. Such surveys also need to capture inequalities in access to these institutions across socio-economic groups.

4.5. Social connections

100. Social connections improve quality of life in a variety of ways. People with more social connections report higher life-evaluations, while many the most pleasurable personal activities involve socialising. The benefits of social connections extend to people's health and to the probability of finding a job, as well as to several characteristics of the neighbourhood where people live (e.g. crime or the performance of local schools). These social connections are sometimes described as 'social capital' to highlight the benefits (direct and indirect) that they bring. As for other types of capital, the externalities stemming from social connections can sometimes be negative: for example, belonging to one group may strengthen in each person a sense of unique identity that breeds a climate of violence and confrontations. This, however, underscores the importance of better analysing the nature of these social connections and the breadth of their effects, rather than undermining their significance. More generally, much of the available evidence suggests that the externalities of social connections are typically positive, not negative.

101. The drivers of changes in people's social connections are not always well understood, and the development of both markets and government policies may have reduced the ties of individuals with their community. What is clear is that a decline in these ties may affect negatively people's life, even when their functions are taken up by market or government alternatives that increase the level of economic activity (as in the case where the informal surveillance of neighbours is replaced by salaried security guards).

102. Research on social connections has traditionally relied on proxy measures such as the number of associations to which each person belongs, or the frequency of activities assumed to result from social connections (e.g. altruistic behaviour and voter turn-out).

However, it is by now accepted that these are not good measures of social connections, and that reliable measures require surveys that inform about peoples' behaviours and activities. In recent years, a number of statistical offices (in the United Kingdom, Australia, Canada, Ireland, the Netherlands and, most recently, the United States) have started surveys measuring various forms of social connections. For example, special modules of the labour force survey in the United States ask people about their civic and political engagement, their membership and volunteering in various organisations, their relationship with neighbours and family members, and how they get information and news. Similar surveys should be implemented elsewhere, based on questions and protocols that allow valid comparisons across countries and over time. Progress should also be made in measuring additional dimensions of social connections (such as trust in others, social isolation, availability of informal support in case of need, engagement in the workplace and in religious activities, friendship across lines of race, religion and social class) building on the experience already accumulated by some countries in these fields.

4.6. Environmental conditions

103. Beyond their importance of sustainability, environment conditions affect the quality of life of people living today in very immediate ways. First, they affect human health both directly (through air and water pollution, hazardous substances and noise) and indirectly (through climate change, transformations in the carbon and the water cycles, biodiversity loss and natural disasters that affect ecosystem's health). Secondly, people benefit from environmental services, such as water and nature, and their rights in this field (including rights to access environmental information) have been increasingly recognized. Third, people value environmental amenities or disamenities and these valuations affect their actual choices (e.g. of where to live). Lastly, environmental conditions may lead to

climatic variations and natural disasters, such as drought and flooding, which affect both the property and the life of affected populations.

104. Measuring the effects of environmental conditions on people's life is, however, complex. These effects manifest themselves over different timescales, and their impacts vary depending on people's characteristics (e.g. where they live and work, their metabolic intake). Further, the strength of these relations is often underestimated because of limits in current scientific understanding and in the extent to which various environmental factors have been subject to systemic investigations.

105. Much progress has been achieved in the last two decades in terms of measuring environmental conditions (building on better environmental data, regular monitoring of indicators, accounting tools and opinion surveys), understanding of their impacts (e.g. evaluation of related morbidity and mortality, labour productivity, economic stakes associated to climate change, biodiversity change, damage of disasters) and establishing a right of access to environmental information. A range of environmental indicators allow measuring the human pressures on the environment, the responses from administrations, enterprises or households to environmental degradation, and the actual state of environmental quality.

106. However, from a quality of life perspective, existing indicators remain limited in important respects: for example, indicators of emissions mainly refer to the aggregate quantities of various pollutants, rather than to the share of people exposed to dangerous doses. Existing indicators should hence be supplemented by regular monitoring of the number of premature deaths from exposure to air pollution; of the number of people lacking access to water services and to nature, or exposed to levels of noise and pollution above dangerous levels; of the damage incurred due to environmental disasters; and of

survey-measures of people's own feeling and evaluations of the environmental conditions of their country and neighbourhood. As many of the effects of environmental conditions on QoL differ across various groups of people, these indicators should refer to people grouped according to various classification criteria.

4.7. Personal insecurity

107. Personal insecurity includes those external factors that put at risk the physical integrity of each person: crimes, accidents, natural disasters are some of the most obvious factors. In their extreme manifestations, these factors can lead to the death of the person involved. While these elements account for only a minority of all deaths, and they are captured by general mortality statistics, the rationale for having specific measures of their frequency is that they have a different emotional effects compared to deaths related to medical conditions, as indicated by the large impacts of bereavement on people's subjective well-being.

108. Less extreme manifestations of personal insecurity such as crime affect quality of life for a significant larger number of people, with even larger number reporting fears of being a victim of a physical aggression. The most remarkable feature of these reports on subjective fears is how little they are related to experienced victimisation: countries with a higher share of people reporting fears of crime do not have a higher frequency of experienced victimisation while, within countries, older and richer people felt more unsafe than younger and poorer people, despite being less likely to be victim of crime.

109. These patterns highlight the importance of developing more regular and reliable measures of personal security to orient public discussions. Victimisation surveys are an essential tool to assess the frequency of crime and the fears it generates. Other tools need

to be mobilised to assess the importance of other threats to personal security, such as domestic violence and violence in countries ravaged by conflicts and wars.

4.8. Economic insecurity

110. Uncertainty in the material conditions that may prevail in the future reflects the existence of a variety of risks, in particular for unemployment, illness, and old age. The realisation of these risks has negative consequences for the quality of life of the person affected, depending on the severity of the shock, its duration, the stigma associated to it, the risk aversion of each individual, and their financial costs.

111. Job loss can lead to economic insecurity when unemployment is recurrent or persistent, the replacement rate is low, and workers have to accept major cuts in pay, hours or both to find a new job. The consequences of job instability are both immediate (as replacement income is typically lower than the earnings on the previous job) and longer term (due to potential losses in wages when the person finds another job). While indicators of these consequences are available, cross-country comparisons are difficult, requiring special investments in this direction. Job insecurity can also be measured by asking workers to either evaluate the security of their present job, or to rate their expectation of losing their job in the near future. Fears of job loss can have negative consequences for the quality of life of each worker (e.g. physical and mental illness, tensions in family life), for firms (e.g. adverse impacts on workers' motivation and productivity, lower identification with corporate objectives) and for society as a whole.

112. Illness can cause economic insecurity both directly, through the medical costs associated to it, and indirectly, through the loss of income due to inability to work. For people with no (or only partial) health insurance, medical costs can be devastating, forcing

them into debt, to sell their home and assets, or to forego treatment at the cost of worse health outcomes in the future. One indicator of economic insecurity due to illness is provided by the share of people without health insurance. However, health insurance can cover different contingencies, and even insured people may incur high out-of-pocket health expenses in the event of illness. To these out-of-pocket health expenses should be added the loss of income that occurs if the person had to stop working and the health (or other types of) insurance does not provide replacement income.

113. Old-age is not a risk *per se*, but it can still imply economic insecurity due to uncertainty on the needs and resources that will be available after having withdrawn from the labour market. Two types of risk, in particular, are important. The first is the risk of inadequate resources during retirement, due to insufficient or volatile pension payments in the future. The second is the risk of volatility in pension payments: while all retirement-income systems are exposed to *some* type of risks, the greater role of the private sector in financing old-age pensions (in the form of both occupational pensions and personal savings) have allowed extending the coverage of pension systems in many countries, but at the costs of shifting risks from government and firms towards individuals, thereby increasing volatility.

114. The many factors shaping economic insecurity are reflected in the variety of approaches used for their measurement. Some approaches try to quantify the frequency of specific risks, while others look at their consequences when a risk materialises, and at the means available to people to protect themselves from these risks (especially those provided by social security programmes). A comprehensive measure of economic insecurity would ideally account for both the frequency of each risk and their consequences, and some attempts in this direction have been made. A further problem is that of aggregating across the various risks that shape economic insecurity, as the

indicators that describe these risks lack a common metric to assess their severity. A further, and more intractable problem, is that of accounting for the long-term consequences of the various policies used to limit economic insecurity on quality of life in the future (through their effects on unemployment and labour force participation).

5. Cross-cutting issues

115. Most of the measurement challenges described above are specific to each feature of quality of life, and the Commission has only hinted at some of the work required, leaving it to future work in each field to detail a concrete action plan. Other challenges are, however, cross-cutting and unlikely to be picked up through initiatives separately undertaken in each field. Three of these issues deserve special attention.

5.1. Assessing links across quality of life dimensions

116. The first cross-cutting challenge is that of better assessing the *relationship* between the various dimensions of quality of life. Some of the most important policy questions for quality of life relate to how developments in one area (e.g. education) affect developments in others (e.g. health status, political voice and social connections), and how developments in all fields are related to those in income. While some of these relations, in particular at the individual level, are poorly measured and inadequately understood, ignoring the cumulative effects of multiple disadvantages will lead to sub-optimal policies: for example, if the loss of quality of life of being both poor and sick far exceed the sum of the two separate effects, governments may need to target more their interventions on those who cumulate these disadvantages.

117. Assessing these links across various dimensions of quality of life will not be easy, as statistical systems continue to be highly segmented across disciplines, with measurement instruments in each field paying only scant attention to developments in other domains. But progress can be achieved by developing information on the ‘joint distribution’ of the most salient features of quality of life (such as hedonic affects, health status, education, political voice) across all residents in a country. While developing this information could be achieved only in the distant future, concrete steps in this direction could be accomplished by including in all surveys a few standard questions that allow classifying respondents based on a limited set of characteristics, and that describe their conditions in a broad range of fields. Investment should also be made in developing longitudinal surveys which could allow controlling for people’s personal characteristics, and better analysing the directions of causation between different factors shaping life.

5.2. Inequalities in quality of life

118. The second cross-cutting challenge is that indicators of quality of life should inform about the *inequalities* in individual conditions in the various dimensions of quality of life rather than just about the *average* conditions in each country. To some extent, the failure to account for these inequalities explains the ‘growing gap’ – identified by the French Presidency when establishing the Commission – between the aggregate statistics that dominate policy discussions and people’s sentiments of their own life.

119. While established methodologies and data sources allow measuring inequalities in the distribution of economic resources in a fairly reliable way, the situation is much less satisfactory with respect to non-monetary dimensions of quality of life. This is especially true as these inequalities cannot be described through information on the *size* of the distribution of these features around their mean. For example, differences in the life span

of people may reflect genetic differences that are randomly distributed in the population: in these circumstances, narrowing the overall distribution of life duration would not make society less 'unequal' in any morally compelling way.

120. The problems, however, go deeper than developing suitable measures. There are as many types of inequalities, and each of them is significant in itself: this suggests that we should avoid the presumption that one of them will always encompass all others. At the same time, because of the links among dimensions, various types of inequalities may strengthen each other. Gender disparities, for example, while pervasive in most countries and groups, are typically much larger for households with lower socio-economic status: the combined effect of gender and socio-economic inequalities is often to exclude young women from poor households from attending school and getting rewarding jobs, denying them possibilities of self-expression and political voice, and exposing them to hazards that put at risk their health. Some of these inequalities (such as those related to class and socio-economic status) have led, over the years, to a wide array of policies and institutions aimed at reducing their intensity and consequences. Others type of inequalities, such as those between various ethnic groups, are more recent (at least in countries that have experienced large waves of immigration) and, arguably, they are set to become more politically salient in the future as immigration continues.

121. It is critical that these inequalities be assessed in a comprehensive way, by looking at differences in quality of life across people, groups and generations. Further, as people can be classified according to different criteria, each with some relevance for people's life, inequalities should be measured and documented for a plurality of groups. Appropriate surveys should be developed to assess the complementarities between the various types of inequalities, and to identify their underlying causes. It is up to the statistical community to regularly feed these analyses with suitable data.

5.3. Aggregating across quality of life dimensions

122. The third cross-cutting challenge to quality of life research is that of *aggregating* the rich array of measures in a more parsimonious way. The issue of aggregation is both specific to each feature of quality of life (as in the case of measures that combine mortality and morbidity in the health field) and more general, requiring the valuation and aggregation of the achievements in various domains of life both at the level of each person and for society as a whole. The search for a scalar measure of quality of life is often perceived as the single most important challenge faced by quality of life research. While this emphasis is partly misplaced – the information content of any aggregate index will always reflect the quality of the measures used in its construction – the demands in this field are strong, and statistical offices do have a role in answering them.

123. Traditionally, the most common response to this demand for parsimony has been that of aggregating a number of indicators (suitably selected and scaled) of *average* performance in various fields at the country-level. The best example of this approach is the *Human Development Index*. This measure has played an important communication role, leading to country-rankings that differ significantly from those based on per capita GDP, especially for some less developed countries. However, all choices on the weights used to construct this (and other similar) index reflect value judgements that have controversial implications: for example, adding the *logarithm* of per capita GDP to the *level* of life expectancy (as done by the *Human Development Index*) implicitly values an additional year of life expectancy in the United States as worth 20 times an additional year of life expectancy in India. More fundamentally, being based on country-averages, these measures ignore the significant correlation between the various features of quality of life

across people, and do not say anything about the distribution of individual conditions in each country. As a result, the combined index will not change if average performance in each domain remains the same while the correlation of individual conditions across domains declines.

124. Several aggregate measures of quality of life are possible, depending on the philosophical perspective taken. Some of these measures are already being used (e.g. average levels of life-satisfaction for a country as a whole). Others could be implemented if national statistical systems made the necessary investment to provide the type of data needed to allow their computation. For example, the U-index, i.e. the proportion of one's time in which the strongest reported feeling is a negative one, requires collecting information on emotional experiences during specific episodes through time-use surveys. Similarly, the methods based on counting the occurrences and severity of various objective features at the individual level (which is linked to the capability approach), before proceeding to construct country-average, require information on the joint distribution of various objective features, while the notion of 'equivalent income' (which is linked to the approach of fair allocations) also require information on individual preferences with respect to these items. Rather than focusing on constructing a single summary measure of quality of life, statistical systems should provide the data required for computing various aggregate measures according to the philosophic perspective of each user.

CHAPTER 3 – SUSTAINABLE DEVELOPMENT AND ENVIRONMENT

1. Introduction

125. Sustainability refers, in a broad sense, to the notion of durability and stability of dynamic processes in the long run. Concern for the sustainability of economic development can be traced back at least as far as Malthus' famous population principle. In a more recent past, this preoccupation has been revived in the early 1970s by the Club of Rome and, soon after, by Nordhaus and Tobin who, in 1974, ventured to build what may be regarded as the first comprehensive measure of sustainable economic development.

126. As popularized by the Brundtland commission in 1987, sustainable development is “development that meets the needs of the present generations without compromising the ability of the future generations to meet their own needs”. This definition implicitly mixes social, environmental and economic components of present and future well-being. Since the Nordhaus-Tobin endeavor, there have been many attempts at building indicators of sustainable development. Some of them directly stem from this Nordhaus-Tobin approach and are therefore familiar to economists or accountants. Some other ones are more specifically environmental and have become very popular amongst NGOs and environmentalists. Several statistical bodies or scholars have, in parallel, developed eclectic approaches that combine several dimensions of sustainability either under the form of extensive dashboards or so-called “composite” indicators.

127. The situation therefore appears as being one of plenty, rather than one of scarcity. But this is not necessarily good news, especially when different indicators provide diverging messages about the sustainabilities of different economic models or about the contributions of different countries to worldwide sustainability. This is a source of

perplexity for the public opinion and for policy makers: which indicators should one focus upon, which are the ones on which empirical investments should be pursued the most actively? In fact, one must admit that, up to now, none of all these existing approaches has been able to receive a level of support that one could consider as sufficiently large.

128. In such a context, the subgroup did not fix to itself the goal of coming out with the new headline indicator able to receive unanimous agreement. Our approach has been more modest but with the hope of being useful to further progress. Starting from the acknowledgement that measuring sustainability is difficult and conflictual, we have considered that our main contribution could be to clarify some of the reasons for this state of affairs. This may sound excessively limited. Yet, if done properly, this clarification exercise can be a powerful lever. It can help escaping some unproductive debates that often result from misunderstandings. It also shows where efforts should concentrate. It can help sorting out what can be done and what is clearly impossible to do. This clarification will clearly leave us far short of proposing a turnkey solution for sustainability assessment, but it can help marking what could be the next steps in this direction.

129. The chapter will be organized as follows. Section 2 will present a review of existing measures. It discusses what we believe to be the pros and cons of the various approaches. All these approaches have their own value and can be of interest to users or policy makers. But we shall nevertheless draw one first strong line of demarcation between very encompassing approaches that, in our view, suffer from their excessive ambition –e.g. summarizing current well-being and its sustainability with only one number-, and approaches that appear more suited for the measurement of sustainability *stricto sensu*. This second group gathers indicators that focus on the twin notions of overconsumption or underinvestment. We shall argue that this second category of approaches all have at least one point in common, which is their focus on stocks of resources that we pass on to future

generations. In this respect, beyond their very large differences, they all can be viewed as particular declinations of a more general framework, known in the literature as the “extended wealth” or “stock -based” approach to sustainability. We shall argue that this framework could provide the basis of a common language favoring exchanges between tenants of different paradigms of sustainability.

130. It is with the help of this general analytical framework that section 3 will try to identify where are these major stumbling blocks in sustainability assessment. We shall start by presenting what this extended wealth or stock-based framework would theoretically allow us to do under ideal conditions of perfect and far-sighted knowledge of expected interaction between the economic and the environmental spheres. From this abstract point of departure, it will then be very easy to enumerate the reasons why this cannot be done in practice, and how this argues in favor of some second best and non mono-dimensional approaches to sustainability, mixing monetary and physical indicators. We shall more particularly emphasize the problems raised by technological and normative uncertainties, and the problems raised by the international dimension of the sustainability issue.

131. The general conclusion (**to be added**) will summarize the main findings and present the incremental proposals that we make for progressing toward a small and well-integrated “micro”-dashboard of sustainability.

2. Taking stock

132. Shortly summarizing the very abundant literature that has been devoted to the measurement of sustainability or durable development is not an easy task. We shall follow

here an imperfect but simple typology that distinguishes (1) large and eclectic dashboards, (2) composite indexes, (3) indexes consisting in correcting GDP in a more or less extensive way and (4) indexes that essentially focus on measuring how far we currently 'overconsume' our resources, this category being itself heterogeneous, since we shall include in it indexes as different as the ecological footprint or adjusted net savings, which, as we shall see, convey very different messages.

2.1. Dashboards or sets of indicators

133. Dashboards or sets of indicators are a widespread approach to the general question of sustainable development. They consist in gathering and ordering series of indicators that bear direct or indirect relationship to socio-economic progress and its durability. In the past couple of decades, international organizations have played a large role in the emergence of sustainability dashboards, with a prominent role for the United Nations. In particular, the 1992 Rio Summit adopted Agenda 21 whose 40th chapter invites signatory countries to develop quantitative information on their actions and realizations.

134. Other international initiatives to build dashboards of sustainable development include those of the OECD and of Eurostat, following the European Council's adoption of its own Strategy for Durable Development (SDD) in 2001. The current version of this dashboard includes 11 indicators at level 1 (covering 10 domains), 33 indicators for level 2 and 78 indicators at level 3, these level 2 and 3 indicators covering 29 sub-themes. Similar national initiatives have accompanied this general movement, albeit in a somewhat scattered way. For instance, the Institut Français de l'Environnement (Ifen) adopted a definition of indicators for the monitoring of the French National Strategy for Durable Development (SNDD) in 2003, with the 11 main indicators adopted for the follow-up of the European SDD. The Ifen maintains in parallel its own larger list of 41 indicators. And

local initiatives have also mushroomed over the last decade, deriving or not from the initial Agenda 21 impulse.

135. For the user, the most striking point is the extreme variety of the indicators proposed. Some are very global ones -GDP growth itself keeps its place, and is even the first indicator in the European Dashboard-, some much more specific -such as the percentage of smokers in the population. Some pertain to outcomes, some others to instruments. Some can easily be related both to development *and* sustainability -literacy performance matters both for current well-being and growth prospects- but other ones either pertain only to current development or only to long run sustainability. There are even some items whose link with both dimensions is disputable or at least of indeterminate sign: is a high fertility level a good thing for sustainability? Maybe yes for the sustainability of pensions, maybe not for environmental sustainability. And is it always the signal of good current economic performance? This probably depends in what we consider as “high” or “low” in terms of fertility?

136. These dashboards are useful in at least two respects. One is the initial step of any analysis of sustainability, which is by nature highly complex and therefore necessitates an effort at establishing the list of relevant variables and encouraging national and international statistical offices to improve measurement of these indicators. The second one is related to the distinction between ‘weak’ and ‘strong’ sustainability. The ‘weak’ approach to sustainability considers that good performance on some dimensions can compensate for low performance along other ones. This allows a global assessment of sustainability with mono-dimensional indexes. The ‘strong’ approach argues that sustainability requires separately maintaining the quantity or quality of many different environmental items. Its follow-up therefore requires large sets of separate statistics, each of them pertaining to one particular subdomain of global sustainability.

137. The drawback remains their heterogeneity, at least in the case of very large and eclectic dashboards, and also some frequent lack of indications about causal links, relationship with sustainability, and/or hierarchies amongst included indicators. Further, as communication instruments, they lack what has made the success of GDP, the power of attraction of a single headline figure that would allow simple comparisons of socio-economic performance over time or across countries.

2.2. Composite indexes

138. Composite indexes are one way to circumvent the problem raised by the extreme richness of dashboards and to synthesize abundant and purportedly relevant information into one single number. The technical report reviews a few of them.

139. For example, Osberg and Sharpe's Index of Economic Well-Being is a composite indicator that simultaneously covers current prosperity (based on measures of consumption), sustainable accumulation, and social topics (reduction in inequalities and protection against "social" risks). Environmental issues are addressed by considering the costs of CO₂ emissions per capita. Consumption flows and wealth accumulation (defined broadly, to include R&D stock, a proxy for human capital, and the costs of CO₂ emissions) are evaluated according to National Accounts methodology. Normalization of each dimension is handled through linear scaling (nine OECD countries) and aggregation relies on equal weights. But the 'green' dimension of this index remains secondary at this stage.

140. Other examples more specifically focus on this green dimension. Such is the case with the 'Environmental Sustainability Index' (ESI) and the 'Environmental Performance Index' (EPI). The ESI covers 5 domains: environmental systems (their global health

status), environmental stress (anthropogenic pressure on these environmental systems), human vulnerability (exposition of inhabitants to environmental disturbances), social and institutional capacity (their capacity to foster effective responses to environmental challenges), and global stewardship (cooperation with other countries in the management of common environmental problems). It is 76 variables that are used to cover these 5 domains. Standard indicators are for instance indicators of air or water quality (e.g. SO₂ and NO_x), health parameters (e.g. infant death rate from respiratory diseases), environmental governance (e.g. local Agenda 21 initiatives per million people)... The EPI is a reduced form of ESI, based on 16 indicators (outcomes), and is more policy oriented: as such, values are scaled according to pre-established policy targets, instead of sample observed data.

141. The messages derived from such an index remain ambiguous. The global ranking makes sense, but is often viewed as presenting developed country's contributions to environmental issues in a too optimistic way. The problem also arises between developed countries. For instance, the index shows a very narrow gap between two countries, the US and France, despite their strong differences in terms of CO₂ emissions. In fact, the index essentially informs us upon a mix of current environmental quality, of pressure on resources and of the intensity of environmental policy, but not about whether a country is on a sustainable path or not: no threshold value can be defined on either side of which we would be able to say that a country is or is not on a sustainable path.

142. In fact, the best use of these indicators is probably to use them as invitations to look more closely at the various components that lie behind them. This kind of function of composite indicators has been often put forward as one of their main *raison d'être*. But this is not sufficient to retain them as measures of sustainability *stricto sensu* that could be placed on the same register as GDP or other accounting concepts. The reason is

twofold. The first is the same as for large dashboards, i.e. the lack of well-defined notion of what sustainability means. The second is the general criticism that is frequently addressed to composite indicators, i.e. the arbitrary character of procedures that are used to weight the various components of the indicator. Such aggregation procedures are sometimes presented as superior to the monetary aggregations that are used for building most of economic indexes, because they are not linked to any form of market valuation. Actually, and we shall come back on this point several times, there are many reasons why market values cannot be trusted when addressing sustainability issues, and more specifically their environmental component. But, monetary or not, an aggregation procedure always means putting relative values on the items that are introduced in the index. In the case of composite sustainability indexes, we have no notion of what argues in favor of putting such or such relative values on all the different variables that matter for sustainability. The problem is not that these weighting procedures are hidden, non-transparent or non replicable: they are often very explicitly presented by the authors of the indexes –and this is one of the strengths of this literature. The problem is rather than their normative implications are seldom made explicit or justified.

2.3. Adjusted GDPs

143. Other candidates for the measurement of sustainability are those that restart from the conventional notion of GDP but try to systematically augment or correct it from elements that standard GDP does not take into account and that matter for sustainability.

144. Nordhaus' and Tobin's sustainable measure of economic welfare (SMEW) may be regarded as the common ancestor to this strand. They provided two indicators. The first one was a Measure of Economic Welfare (MEW) obtained by subtracting from total private consumption a number of components that do not contribute positively to

welfare (such as commuting or legal services) and by adding monetary estimates of activities that contribute positively to welfare (such as leisure or work at home). The second step consisted in converting the MEW into the SMEW by taking into account changes in total wealth. The SMEW measures the level of MEW that is compatible with preserving the capital stock. To convert the MEW into the SMEW, Nordhaus and Tobin used an estimate of total public and private wealth including reproducible capital, non reproducible capital (limited to land and net foreign assets), educational capital (based on the cumulated cost of years spent into education by people belonging to the labor force) and health capital, based on a method of permanent inventory with a depreciation rate of 20% per year. But they did not in the end include estimates of environmental damages or natural resource depletion.

145. Two strands have developed from this seminal contribution. The first one has tried to enrich Nordhaus and Tobin's approach, sometimes deviating in an increasing fashion from the criterion of accounting consistency. One example is the ISEW (Index of Sustainable Economic Welfare) and the GPI (Genuine Progress Indicator). These indicators deduce the costs of water, of air and of noise pollutions from consumption and accounts for losses of wetlands, farmland, primary forests, for other natural resource depletion, and for CO₂ damages and ozone depletion. The evaluation of natural resources depletion is measured as the investment necessary to generate a perpetual equivalent stream of renewable substitutes.

146. In all countries for which they are available, ISEW and GPI are very close and at some point in time start diverging from GDP. This decorrelation has led some authors to put forward a so-called 'threshold' hypothesis, according to which GDP and welfare move in the same direction up to a certain point, beyond which the continuation of GDP growth does not allow any improvement in well-being anymore -and eventually a decline-. In

other words, according to such indicators, sustainability is already far behind us and we have already entered a phase of decline.

147. The other strand is more firmly integrated into the realm of national accounting. It is based on the so-called SEEA (System of Environmental Economic Accounting), a satellite account of the SNA (Standard National Accounts). The SEEA brings together economic and environmental information in a common framework to measure the contribution of the environment to the economy and the impact of the economy on the environment. The UN Committee of Experts on Environmental-Economic Accounting (UNCEEAA), created in 2005, is now looking forward to mainstream environmental-economic accounting, elevate the SEEA to an international statistical standard by 2010 and advance the SEEA implementation in countries.

148. The SEEA comprises four categories of accounts. The first considers purely physical data relating to flows of materials (materials drawn into the economy and residuals produced as waste) and energy and marshals them as far as possible according to the accounting structure of the SNA. The second category of accounts takes those elements of the existing SNA which are relevant to the good management of the environment and make the environment-related transactions more explicit. The third category of accounts comprises accounts for environmental assets measured in physical and monetary terms (timber stock accounts for instance).

149. These first three aspects of the SEEA are of a crucial importance as building blocks for any form of sustainability indicators. But what is at stake here is the fourth and last category of SEEA accounts that considers how the existing SNA might be adjusted to account (exclusively in monetary terms) for the impact of the economy on the environment. Three sorts of adjustments are considered: those relating to depletion, those

concerning so-called defensive expenditures (protection expenditures being the most emblematic ones), and those relating to degradation.

150. It is these environmental adjustments to existing SNA aggregates that are better known under the rather loose expression of 'Green GDP'. It is close, in spirit, to the computation of a 'Hicksian' income concept that includes natural capital. Indeed, just as GDP (Gross) is turned into NDP (Net) by accounting for consumption of fixed capital (depreciation of produced capital), the idea is that it would be meaningful to compute an eaNDP (environmentally-adjusted) taking into account the consumption of natural capital. The latter would comprise resource depletion (the over-use of environmental assets as inputs to a production process) and environmental degradation (the value of the decline in the quality of a resource, roughly speaking).

151. Green GDP or eaNDP remain however the most controversial outcomes of the SEEA, and as such the less implemented by statistical offices, because of the many problems that are raised by these two concepts.

152. Valuing environmental inputs into the economic system is the (relatively) easier step. Since these inputs are incorporated into products which are sold in the market place, it is possible (in principle) to use direct means to assign a value for them based on market principles. On the contrary, as residuals (leading to degradation) are outputs, there is no direct way to assign a value to them. All the indirect methods of valuation will depend to some extent or another on 'what if' scenarios. Thus, translating valuations of degradation into adjustments to macro-economic aggregates takes us beyond the realm of *ex-post accounting* into a much more hypothetical situation. The very speculative nature of this sort of accounting explains the great discomfort and strong resistance in this area for many accountants.

153. Experience from the field suggests two main options for valuing degradation. The first of these relies on damage-based estimates and the second relies on cost-based estimates. Indeed, the damage-based option answers the question ‘how much damage is caused by environmental degradation?’ and tries to estimate the loss of welfare caused by the effect of residual generation on human health and thus on human capital.
154. Cost-based estimates instead answer the question of ‘how much would it cost to avoid environmental degradation?’, and in turn can be divided into two types. The first type of cost-based estimates relies on *maintenance costs*, that is the value of the costs which would have had to be incurred to remedy the environmental degradation caused by current production and consumption, and leads to ‘environmentally adjusted’ aggregates for those costs. As such, it estimates what the accounting entries would have been for the same level (and structure) of activities and demand if all the costs associated with environmental degradation had been incurred and internalized within market prices. The problem with this approach is that the resultant price rises involved (potentially high for non-marginal changes) are likely to bring about a change in behavior, which would affect the level of demand for those products (and thus the level of output and/or the choice of the technology of production).
155. The second type of cost-based estimates attempts to overcome those limitations and rather answer the following question ‘What level of GDP would be achieved if producers and consumers faced a different set of relative prices in the economy due to the existence of actual prices for environmental functions?’ It is thus a forward-looking modeling approach (known as greened economy modeling) rather than a one-time adjustment to a selected number of macro-aggregates and interest then focuses less on the new ‘greened’ aggregates themselves than on the gap between the existing economy and the ‘greened’ version (and possible transition paths between the two).

156. This need to rely on some joint physical and economic modeling at some stage or another will systematically reappear later on, whatever the approach that one wishes to retain. But there is a more fundamental problem with green GDP, that applies as well to Nordhaus and Tobin's SMEW or to the ISEW/GNI indexes. None of these measures characterize sustainability *per se*. They measure what hypothetical level of well-being or consumptions would be sustainable in the long run. This is only one part of the answer to the question of sustainability. What we ultimately need is an assessment of how far we are from these sustainable targets. In other words, what is needed are rather measures of **overconsumption** or, to put in dual terms, of **underinvestment**. This is precisely what our last category of indicators purports to do.

2.4. Indexes focusing on overconsumption or underinvestment

157. Under this heading, we regroup all kinds of indicators that address this question of sustainability in such terms of overconsumption, underinvestment or excessive pressure on resources. As for GDP or other aggregates, trying to do it with single numbers requires the choice of a metric and an explicit aggregation procedure. Though such indicators tend to be presented in flow terms, they are built upon the assumption that there corresponds, to the measured flows, some stocks that are relevant for sustainability, i.e. that are being transmitted to future generations and determine their opportunity sets.

2.4.1. Adjusted net savings (ANS)

158. Adjusted net savings (also known as genuine savings) is a sustainability indicator building on the concepts of green national accounts. The theoretical background is here again the standard Hicksian concepts of income and saving, both of which refer to the maintenance of a constant stock of 'wealth': the 'genuine' economic income of a

generation is defined as the flow that can be consumed without depleting the existing stock of wealth; similarly, 'genuine saving' is taken to be the amount of variation of total wealth over a given time period, such as the year. Alternatively one may prefer to look at 'genuine investment', referring to the change in the total capital stock, insofar as what is being measured in practice under the name of 'extended wealth' is an 'extended capital stock'. Such concepts clearly appear to be the relevant economic counterparts of the notion of sustainability, in that they include not only the natural resource part of the problem but also (in principle at least) those other ingredients necessary to provide future generations an opportunity set that is at least as large as that being currently available to living generations.

159. Equipped with these definitions, researchers at the World Bank took the lead in computing 'genuine savings' or 'adjusted net savings' (ANS) for a large number of countries. The World Bank Report for 2004 contains estimates of these 'net domestic saving' for almost all countries in the world. The concept of 'extended wealth' (EW) chosen by the authors is not limited to natural resources, but also includes physical, productive capital, as measured in traditional national accounts, and human capital.

160. Indeed, adjusted net savings measure the 'true' rate of savings in an economy after taking into account investments in human capital, depletion of natural resources (forest, energy, mineral) and damage caused by pollution (CO₂ emissions). Empirically, adjusted net savings are derived from standard national accounting measures of **gross national savings** by making four types of adjustments. First, estimates of capital consumption of produced assets are deducted to obtain **net national savings**. Then current expenditures on education are added to net domestic savings as an appropriate value of **investments in human capital** (in standard national accounting these expenditures are treated as consumption). Next, estimates of the **depletion of a variety of natural resources** are

deducted to reflect the decline in asset values associated with their extraction and harvest. Estimates of resource depletion are based on the calculation of resource rents. An economic rent represents the excess return to a given factor of production. Rents are derived by taking the difference between world prices and the average unit extraction or harvest costs (including a 'normal' return on capital). Finally, **global pollution damages from carbon dioxide emissions** are deducted². Negative adjusted net saving rates imply that 'extended wealth' is in decline, and as such provide a warning of unsustainability.

161. How does this indicator compare with standard measures of saving and investment in national accounts? World Bank-computed ANS for developed countries such as France and the US show the dynamics is almost exclusively driven by gross savings, while the gap in levels between ANS and gross savings is mostly due to capital consumption and human capital accumulation, as evaluated by education expenditures, whereas natural capital changes only play a relatively marginal role. Moreover, they show that most developed countries are on a sustainable path, while many emerging or developing countries are not. In particular, natural resources-exporting countries tend to be unsustainable, according to this measure.

162. Such an approach makes sense to many economists as it benefits from a clear and sound theoretical framework defining sustainability, but current methodology underlying empirical calculations of the ANS per country has well-known shortcomings, articulated by the authors themselves. In general, the relevance of the ANS approach crucially depends on **what** is counted (the different forms of capital passed on to future generations), namely what is included in the 'extended wealth', and on **the price** used to count and aggregate in a context of imperfect or indeed non-existing valuation by markets

² As for local pollution damages, they are difficult to estimate without location-specific data. Nevertheless, an augmented version of ANS for local pollution is also provided by taking into account health damages due to urban air pollution (particulate matter PM10).

–the problem that we already mentioned when discussing the implicit prices used by composite indexes.

163. Indeed, a major shortcoming of the ANS produced by the World Bank is the fact that the adjustment for environmental degradation is limited to global pollution damages from carbon dioxide emissions. The authors acknowledge that the calculations are not comprehensive in that they do not include some important sources of environmental degradation such as underground water depletion, unsustainable fisheries, and soil degradation, and *a fortiori* biodiversity loss.

164. For those natural assets that are taken into account, pricing techniques remain the major issue. For exhaustible resources, the World Bank's estimates of ANS rely on current prices. In theory, the use of market prices to evaluate flows and stocks is only warranted in a context of perfect markets, which is clearly not the case in reality, and in particular not for natural resources, where externalities and uncertainties are paramount. In particular, market prices for fossil energy sources and other minerals have tended, in recent years, to fluctuate widely, causing significant swings in measures of ANS based on current market prices.

165. As for pricing environmental degradations, things turn out to be even trickier because of the absence of any market valuation that could be used as a starting point: in theory, accounting prices must be evaluated by modeling long term consequences of given changes in environmental capital and how they impact prospective well-being. But practical implementation raises considerable problems. Under the current state of the art, prices used to value carbon emissions in existing estimates of the ANS are not able to give it any significant role in the global assessment of sustainability, and this casts doubts on the usefulness of the indicator as a guide for environmental policy.

166. Finally, while computing ANS per country, we leave out the very **international nature of sustainability**. Indeed, one may feel uneasy in front of the message conveyed by the ANS regarding resources (e.g. oil) exporting countries. In such exporting countries, unsustainability only comes from an insufficient rate of reinvestment of the rents generated by the exploitation of the natural resource, and “over-consumption” by importing countries is not an issue at all. Developed countries, generally less endowed in natural resources but richer in human and physical capital than developing ones, would then unduly appear sustainable. As a consequence, a case has been made by some authors for imputing consumption of exhaustible resources to their final consumers, i.e. the importing countries.

167. To be exact, if scarcities were fully reflected in the prices at which these exhaustible resources are sold on international markets, it is true that there would be no reason for doing such a correction. However, when prices are non competitive, the importing country is able to pay its imports less than would be required for efficiency, it will have a responsibility in global non-sustainability that is not captured by the money-value of its imports. Low prices allow this country to overconsume and to transfer the long-term costs of this over-consumption to the exporting country.

2.4.2. Footprints

168. Although apparently quite different from ‘extended wealth’ notions, various attempts at measuring sustainability through the use of ‘footprints’ are, in fact, also inspired by the general approach of comparing current flows of consumption and their induced effects on some dimensions of the environment with an existing stock. In this sense, they may also be regarded as ‘wealth’ measures in which the focus is exclusively on natural capital, and the valuation convention differs from the ANS one, in that no market prices are explicitly

used. However, only insofar as natural capital does aggregate different varieties of environmental goods, can footprints be regarded as ‘physical’ measures’.

169. The Ecological Footprint (EF hereafter) measures how much of the regenerative capacity of the biosphere is used by human activities (consumption). It does so by calculating the amount of biologically productive land and water area required to support a given population at its current level of consumption and resource. A country’s Footprint (demand side) is the total area required to produce the food, fiber and timber that it consumes, absorb the waste it generates, and provide areas for its infrastructures (built-up areas). On the supply side, biocapacity is the productive capacity of the biosphere and its ability to provide a flux of biological resources and services useful to humanity.

170. Results are well-known and rather striking: since the mid-1980s humanity’s footprint is larger than the planet's carrying capacity and in 2003 humanity’s total Footprint exceeded the Earth’s biocapacity by approximately 25 per cent (we would have needed an extra 25% Planet to meet our needs, to say it informally). While 1.8 global hectares per person are available world-wide, Europeans use 4.9 global hectares per person and North Americans use twice that amount, that is much more than the actual bio-capacities of the two geographical zones.

171. As such, this index may be regarded as an extended account approach even if its results are not expressed in monetary terms. Indeed, this indicator shares with accounting approaches the idea of reducing heterogeneous elements to one common measurement unit (the global hectare, e.g one hectare with productivity equal to the average productivity of the 11.2 billion bioproductive hectares on Earth). It assumes at least the substitutability of different forms of natural capital as it assumes different natural capital goods are additive in terms of land area, but strongly stands against weak sustainability

assumptions. In fact, there is no role at all for any saving or accumulation virtue: any positive ecological surplus (biocapacity exceeding ecological footprint) does not entail an increase in some natural capital stock and hence an improvement of the productive capacity in the future. Similarly, non-renewable resource depletion (e.g. oil) and induced threats to sustainable growth based on such resources are rather treated from the waste assimilation (CO₂ emissions) point of view than from a depletion dynamic analysis. A fortiori, saving and accumulating manufactured capital does not help sustainability.

172. Results are also problematic for measuring a country's own sustainability, because of the substantial anti-trade bias inherent of the Ecological Footprint methodology. The fact that densely populated regions, hence low biocapacity countries like the Netherlands, have ecological deficits whilst sparsely populated regions, hence high biocapacity countries such as Finland, enjoy surplus can be seen as part of the normal trade of goods to the mutual benefits of both, and is by no means an indicator of unsustainability. Indeed, the current tendency is to move away from comparing a country's EF with its *own* biocapacity, and proposed rather to divide all countries' EFs by *global* biocapacity. By doing so, one acknowledges that EFs are not intended as measures of a country's own sustainability, but more of its contribution to global unsustainability.

173. Overall, this means that EF could be at best an indicator of instantaneous unsustainability *at the worldwide level*. EFs for countries should rather be used as an indicator of inequality in the exploitation of natural resources and interdependencies between geographical areas. Moreover, even the worldwide ecological deficit emphasized by the Ecological Footprint may not convey the message it pretends. Indeed, one can show that worldwide unbalance is mostly driven by CO₂ emissions, expressed in hectares of forest needed for storage. By definition, worldwide demand placed on cropland, built-up land and pasture can not overcome world biocapacity.

174. As a result, less encompassing footprints, but more firmly defined, such as the ‘Carbon footprint’ would seem better suited to using alongside synthetic indicators, insofar as they are more clearly physical measures of stocks, not relying on specific assumptions about productivity or equivalence factor. As far as communication is concerned, such an indicator has exactly the same capacity to send strong messages in terms of over-utilization of the absorbing capacity of our planet. And it also has the interesting feature of being computable at any level of disaggregation. This makes it a powerful instrument for monitoring behaviors of individual actors.

3. Summarizing sustainability with a single number: why is it so difficult ?

175. Let’s summarize the main messages. The previous section has shown the large number of existing attempts to quantify sustainability. Such abundance is a serious drawback insofar as different synthetic indicators convey widely divergent messages. This leads to strong perplexity for the statistician or the policy maker. It urges to return to the fundamental questions: What do we want to measure exactly? What would be needed to do it in a satisfactory way? What are the second best strategies to adopt when these ideal conditions are not met?

3.1. What do we want to measure ?

176. What do we want to measure? Since the Brundtland Report, the notion of sustainable development has expanded to become an encompassing concept absorbing all dimensions of present and future economic, social and environmental well-being. Such an ambition is justified, but it covers all the domains considered by the three subgroups of the commission. The mandate of our environment/sustainability subgroup was narrower than that: it concentrated on the ‘durability’ component of ‘durable development’. This

question of durability can be expressed in the following terms: assuming we have been able to assess what is the current level of well-being, the question is to know whether the continuation of present trends allows or does not allow the preservation of this current level of well-being.

177. It seems actually sound to separate these two notions of current well-being and of its durability or sustainability, because the two questions are interesting for themselves. This provides a first guide for sorting out the many different approaches reviewed in the first half of this chapter.

178. Extensive dashboards of durable development reviewed in section 2.1 are the method for which this interpenetration between measurement of current well-being and its sustainability was at its highest. This is not to say that dashboard approaches are to be excluded. Quite the contrary: our final conclusion will be that a monodimensional view of sustainability remains certainly out of reach. But we want to end up at most with a limited number of indicators –a ‘micro’ dashboard- and one that is specifically dedicated to the sustainability issue, based on a clear notion of what sustainability means.

179. Composite indicators based on such eclectic dashboards raise similar problems, with the additional one that the way in which their various items are weighted is always arbitrary with consequences that are seldom made explicit.

180. Measures of a **sustainable standard of living**, such as the Green GDP are also insufficient for assessing sustainability: measuring the sustainability of a society’s consumption requires a comparison between its actual consumption and its ‘genuine’ production, or in more comprehensive terms, a comparison between its actual and sustainable levels of well-being. In addition, the proximity that such a sustainability indicator would necessarily have with standard GDP could be a source of confusion. If

there are two GDP indicators, which one should we use in which context? Which conclusion would we draw from the fact that such or such country's green GDP is x% or y% of its GDP defined in standard terms? Does this necessarily imply that this country is running an unsustainable path?

181. All this makes the appropriate sustainability index more akin to a concept of net investment or disinvestment and this is precisely the route of which the ANS was a particular example, but which is also implicitly followed by footprint indicators more specifically focused on the renewal or decumulation of environmental assets. The argument goes as follows: the capacity of future generations to have standards of well-being at least equal to ours depends upon the fact that we pass on to them sufficient amounts of all these assets that matter for well-being. Let's note W the 'extended wealth' index that would be used to quantify this stock of resources. Measuring sustainability amounts to testing whether this global stock or some of its components evolve positively or negatively, i.e. computing its or their current rates of change dW or dW_i . If negative, this means that downward adjustments in consumption or well-being will be required sooner or later. This is exactly what one should understand by 'non-sustainability'.

182. In our view, such a formulation of the sustainability issue has a strong potential for providing the common language necessary for constructive debates between actors coming from very different fields. Just to take one example, it fully answers to one of the longstanding objections made to GDP by environmentalists, i.e. the fact that ecological catastrophes can increase GDP through their implied impact on economic activity. In an extended wealth approach, an ecological catastrophe is registered as a destruction of capital. This accounts for the fact that it deteriorates sustainability by decreasing the resources available for generating well-being tomorrow. This can be avoided only if some

action is taken to repair the corresponding damages, these actions being counted as positive investment. index.

3.2. What would be the requirements of an appropriate index ?

183. Now, we have seen that both ANS and footprints evaluations are subject to many objections and can be considered, at best, as a proxies of what would be genuine indexes of change in extended wealth. Returning to fundamentals means asking precisely what would be required to measure such a dW indexes in a satisfactory way. Assuming away the measurement problems at first, we have to be more specific about several concepts: what is to be sustained?, how do the various assets that will be passed on to future generations affect this measure of well-being?, and how should they weighted against each other?

184. It is clearly this latter question that is the more problematic and on which oppositions between tenants of monetary and physical indicators tend to cristallize. Is there actually some reasonable perspective for monetarizing everything, or should we consider that this is possible only up to a certain point?

185. If all assets were traded on perfect markets by perfectly forward-looking agents, one could defend the position that their current prices reflect the discounted streams of their future contributions to future well-being. But many assets are not traded at all, and even for those that are traded on a market, it is unlikely that current prices fully reflect this dimension, due to market imperfections, myopia and uncertainty. This implies that a true measure of sustainability requires a dW index in which assets are not valued at market prices, but rather on the basis of imputed 'accounting prices' based on some objective physical or economic modeling of how future damages to the environment will affect

well-being, exactly as it requires an exact evaluation of how current additions to the stock of human or physical capital are likely to improve or help maintain well-being in the future.

186. Some recent theoretical literature has clarified the requisites of such an exercise. One is a full economic and physical projection of how initial conditions determine the future joint path of economic, social and environmental variables. And the other is the *a priori* definition of how this path translates in terms of well-being at all future dates, i.e. the knowledge of the social utility function, generally formalized as a discounted sum of well-being over all future periods.

187. Equipped with such instruments, one would be able to derive sustainability indexes that have the properties that one should expect, i.e. a capacity to anticipate future declines of well-being below its current level. Some simulations proposed in the technical report illustrate some aspects of this capacity. First of all, this sustainability index is the best suited for sending correct forewarnings to countries who are on unsustainable paths because of an insufficient rate of accumulation or renewal of their produced capital --be it human or physical. And this is of course an important property: even if environmental issues are of a considerable importance, we cannot ignore these other dimensions if sustainability.

188. Second, such an indicator is potentially able to encompass the 'strong' view of unsustainability, i.e. problems arising from the depreciation of environmental assets that are essential to human well-being or even survival. It is when it relies on fixed price levels for natural and non-natural assets that he cannot capture this form of unsustainability. But if we were able to derive this index from a physico-economic model predicting future interactions between the economy and the environment in a reliable way, then this index

would send us correct forewarnings of non sustainability, through strong upward curbing of the relative “imputed” price of these critical natural assets.

189. But all the problem is with those “ifs”. This construction remains fully theoretical. It shows us at best the direction in which index builders could try to go. Many obstacles that exist to its implementation. These obstacles deserve detailed examination.

3.3. Technological uncertainties argue in favor of a more hybrid approach

190. Measuring sustainability with a single dW index works well only under two strong assumptions: one is a perfect prediction of future eco-environmental developments; the second one is a perfect knowledge of how these developments are going to affect well-being. These two assumptions are clearly at odds with our real world situation. Debates on eco-environmental perspectives are dominated by high levels of ignorance or uncertainty concerning future interactions between the two spheres, and by lack of consensus upon the definition of the objective function itself.

191. Let’s briefly develop the first point. The future is fundamentally uncertain. This means that indicators can be at best interpreted in probabilistic terms: they can give no more than a likelihood that we are or are not on an unsustainable path, with the two symmetric risks of warning us unnecessarily of a future unsustainability that will not materialize or, on the other hand, letting us believe that we are on a sustainable path while we are not. Uncertainty takes many forms, some of them amenable to probability computation, but some other ones much more radical. It affects not only the parameters of models that one may try to use to project eco-environmental interactions, but also the structure of these models themselves, the measurement of current stocks, and even the list of these natural assets whose stocks and future evolutions have to be taken into account. It has been recognized for long that most of the debates concerning long term environmental

changes are not necessarily ideological ones but reflect different beliefs on future eco-environmental scenarios. There was no reason why sustainability measurement could escape such a difficulty.

192. To this problem many solutions might be considered. One is to do what every prospectivist does when he wants to emphasize the uncertain nature of future trends, i.e. work with scenarios or provide confidence intervals. Our discussion of the ANS has shown for instance the incidence of re-computing its values for different pricings of CO₂ emissions. This mode of presentation could be adopted systematically.

193. But it can remain insufficient, and we have seen that such seems to be the case with CO₂ emissions: large variations in their valuation do not avoid having their impact largely offset by the positive valuation accumulation of other capital goods. This can be an argument in favor of a separate accounting of this and other forms of environmental degradation, and drives us back to the usual distinction between weak and strong sustainability. The point is not that aggregate dW indexes are intrinsically unable to account for situations of strong unsustainability. The point is that we would be able to do so only by adopting extreme valuations of critical environmental assets and that we are not that well equipped for exactly quantifying what should be these extreme valuations.

194. In such cases, and *a fortiori* for items on which we do not have a single guesstimate of a monetary value, a separate physical accounting is unavoidable. The problem is then to have it presented in a suggestive ways. Monetary indexes have this advantage that they use units that talk to everyone. In addition, they can be put in relation to other monetary quantities: this is what we do when we compute extended savings rates, and the orders of magnitude of such savings rates can be understood easily. On the other hand, tons of CO₂ emissions are not a very informative numbers if we do not have some reference of how

many tons can be emitted each year without severe future climatic consequences. Finding more suggestive ways to highlight such a figure is necessary, if we want the indicator to have some impact in the debate. It has been the major success of the ecological footprint to be able to express pressure on the environment in a unit that is easily understandable. The EF indicator has limits that makes it problematic to many observers. But the general idea to use the footprint as a generic unit for the different forms pressures than men exert on earth's regenerative capacity is an option to be considered –such a metric is used, for instance, with the more focused concept of Carbon Footprint.

3.4. Uncertainty is also normative

195. The second idea that there exists some normative uncertainty is less often emphasized in the eco-environmental literature but probably of equal importance. It is that there can be as many indices of sustainability as there are definitions of what we want to sustain, and this poses a second limit to the practical choice of a single headline dW indicator. In standard National Accounting practice, the normative issue of defining preferences is generally avoided through the assumption that observed prices reveal the real preferences of people. No explicit normative choice is therefore to be made by the statistician. But as soon as we consider that market prices cannot be trusted, alternative imputed prices must be computed, whose values will strongly depend upon normative choices.

196. Can we solve this normative problem? One could try to solve it empirically, trying to infer the definition of well-being from current observations of how people value environmental factors compared to economic ones, using contingent valuations or direct measures of the impact of environmental amenities on indexes of subjective well-being such as those considered in subgroup 3. But can contingent evaluations or subjective

measures established today in a certain eco-environmental setting be used to predict what will be the valuations of future generations in eco-environmental settings that may have become very different? Some may argue that our descendants may become very sensitive to the relative scarcity of some environmental goods to which we pay little attention today because they are still relatively abundant, and that the precautionary principle should therefore require that we immediately place a high value on these items just because we think that our descendants may wish to do so.

197. Another aspect of this normative question is to know how sustainability indexes should aggregate individual preferences. This depends on how distributional considerations are taken into account in our measures of current well-being. For instance, if we consider that the headline indicator of current well-being must be the total disposable income of the bottom 80% of the population, or of the bottom 50%, rather than global disposable income, then indicators of sustainability must be adapted to such an objective function. This would be completely in line with what was one of the other aspects of the Brundtland definition of sustainability that is often by-passed, i.e. its additional concern for the distribution of resources *within* as well as *between* generations. In a world with natural tendencies to increasing inequalities within countries, messages concerning sustainability will differ depending on the goal that we fix to ourselves. A specific attention to distributional issues may even suggest enlarging the list of capital goods that matter for sustainability: the ‘sustainability’ of well-being for the bottom x% of the population may imply some specific investment in institutions that help protecting efficiently this population from poverty. In principle, the theoretical framework of dW approaches tells us how we could ideally put some value on such ‘institutional’ investment. But needless to say that the prospect of being able to do so is still more remote than for other assets.

3.5. An additional source of complexity: the transnational dimension

198. Additional problems arise concerning the properties of sustainability indicators in a multinational context. Advocates of the ANS defend the message that unsustainability problems generally concentrate in poor resource-exporting countries even if it is in developed countries that these resources are ultimately consumed. The argument is that, if markets work properly, the pressure that developed countries exert on other countries' resources is already reflected in the prices that they pay for importing these resources. If, despite this cost of their imports, they are still able to have a positive ANS, this means that they invest enough to compensate for their consumption of natural resources. It is then the responsibility of exporting countries to reinvest the income from their exports in sufficient quantities if they want to be also on a sustainable path.

199. Yet this is true only under the assumption of efficient markets. If markets are not efficient and if the natural resource is underpriced, then the importing countries benefit from an implicit subsidy and the exporting ones are taxed. This means that effective sustainability of the former is overestimated, while the sustainability of the latter is underestimated. And this problem will be all the more crucial when there are no markets at all, or in the presence of strong externalities.

200. To illustrate this issue in a very crude way, let's imagine a very simple two-country setting, where both countries produce and consume with external effects on the stock of a natural resource that is a global public good with free access, treated as an exhaustible resource. Country 2 uses a clean technology that has no impact on the natural resource, while country 1 uses a 'dirty' one that leads to a depreciation of this resource. Let's push further the asymmetry by assuming that it is only country 2 that is affected by this degradation of the environmental good. Country 1 is completely indifferent to the level of

this environmental good, for instance because its geographical characteristics fully protects it from consequences of its degradation.

201. With such a setting it is natural to redefine countries 1 and 2 as being respectively ‘the polluter’ and ‘the polluted’. In this setting, there are two ways to consider sustainability. One possibility consists in computing changes in extended wealth for each country using country-specific accounting prices for the natural resource. The idea is that the environmental good is a common asset, but valued differently by each country, because they are not concerned in the same way by its degradation. In this example, such an accounting price for the polluter will be zero because we have assumed that it is not impacted at all by environmental changes, and this implies that he attributes no value at all to the environmental asset. On the other hand, the polluted gives a positive value to this asset. The message from such a way to apply the extended wealth concept is that the polluter is on a sustainable path, while the polluted is not.

202. From a certain point of view, this result is not absurd. It is formally correct to say that the polluter is not confronted to the perspective of a declining well-being, contrarily to the polluted. But from another point of view, the message to policymakers is clearly misleading. The polluted can do nothing to restore its sustainability. It is only a change in the polluter’s technology that can help restore sustainability for this polluted country. We are in need of indexes that would send such a message. The popularity of footprint indicators precisely stems from the fact that -whatever their other limitations- they are able to send such messages to policy makers or public opinion.

203. In an extended wealth framework, the solution to that issue can be to try to develop them in two versions of the index: one devoted to the measurement of country-specific sustainabilities, and the other one to measuring contributions to changes in the worldwide

stock of extended wealth, i.e. national contributions to overall sustainability or unsustainability. This duality more or less reproduces the two different points of view of the standard ANS and of the ecological footprint. Yet the way to do it exactly needs elaboration. In particular, it raises the question of knowing how we choose to aggregate levels of well-being for the different nations, a problem that is the trans-national version of the distributive question raised for the intranational level at the end of section 3.2. Does worldwide sustainability mean the preservation of well-being for the average of the world population, or for those of its members the more directly confronted to the adverse consequences of such or such environmental threat. We know that the issue of global warming is a good example of that question, with a strongly uneven distribution of consequences of climatic change.